DOCUMENT RESUME

ED 478 708 PS 031 428

AUTHOR	Wilkinson, Ian A.G.; Hattie, John A.; Parr, Judy M.; Townsend, Michael A.R.; Fung, Irene; Ussher, Charlotte; Thrupp, Martin; Lauder, Hugh; Robinson, Tony
TITLE	Influence of Peer Effects on Learning Outcomes: A Review of the Literature.
SPONS AGENCY	Ministry of Education, Wellington (New Zealand).
PUB DATE	2000-06-00
NOTE	182p.; Produced by Auckland UniServices Limited.
AVAILABLE FROM	Auckland UniServices Limited, %The University of Auckland, Private Bag 92019, Auckland 1, New Zealand. Tel: 64-9-373-7522; Fax: 64-9-373-7412; e-mail: uniservices@auckland.ac.nz; Web site: http://www.uniservices.auckland.ac.nz.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC08 Plus Postage.
DESCRIPTORS	Educational Research; *Elementary School Students; Family Environment; *Influences; *Learning Processes; Models; *Peer Influence; Peer Relationship; School Organization; *Secondary School Students; Teaching Methods

ABSTRACT

This report presents a literature review and conceptual model summarizing the influence of peer effects on learning outcomes. The report describes the approach to the review and provides a theoretical account of the environments, mechanisms, and processes that mediate learning among peers. It then summarizes the literature on compositional effects at each level of school organization--groups, class, and school--and suggests how these effects might implicate peers by making connections to the theoretical account of peer-mediated learning. Next, the report makes linkages across different levels of inquiry in order to develop a conceptual model of peer influences on learning. A multi-layer model is proposed, with effects propagating from school-level influences to class-level influences to group-level influences to ambient and configured environments for learning among peers. It is proposed that the bulk of the effects are indirect; hence, peer effects "look" smaller the further one moves away from the instructional coalface because they are mediate by intervening layers. It is noted that there may also be reciprocal effects whereby peers influence teachers and school organization and management, although the magnitude of these effects is undetermined. In concert with the three layers of influence, it is argued that family resources have greater effects at uppers layers and smaller effects at the lower layers; conversely, curriculum and teaching resources have greater effects at lower layers and smaller effects at upper layers. Home and school supports for learning carry the lion's share of the weight in predicting student learning outcomes, whereas peer effects, as currently constituted, carry much less weight. Finally, the report describes four instructional approaches that utilize peer resources to maximize learning. These models demonstrate additional ways of capitalizing on peer effects beyond altering student composition. (Contains 637 references.) (Author/HTH)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 708

UniServices
Auckland UniServices Limited

Influence of Peer Effects on Learning Outcomes:

A Review of the Literature

AUCKLAND UNISERVICES LIMITED

a wholly owned company of

THE UNIVERSITY OF AUCKLAND

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Ian A.G. Wilkinson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Prepared for:

Ministry of Education
45-47 Pipitea Street
Thorndon
Wellington

Date: June 2000

Prepared by:

Ian A.G. Wilkinson, John A. Hattie, Judy M. Parr, Michael A.R. Townsend (with assistance from Irene Fung and Charlotte Ussher)
The University of Auckland
Martin Thrupp
University of Waikato
Hugh Lauder & Tony Robinson
University of Bath

PS 031428

BEST COPY AVAILABLE

Reports from Auckland UniServices Limited should only be used for the purposes for which they were commissioned. If it is proposed to use a report prepared by Auckland UniServices Limited for a different purpose or in a different context from that intended at the time of commissioning the work, then UniServices should be consulted to verify whether the report is being correctly interpreted. In particular it is requested that, where quoted, conclusions given in UniServices reports should be stated in full.

Ian A. G. Wilkinson
Now at The Ohio State University

This research was funded by the Ministry of Education, Wellington, New Zealand. We wish to acknowledge the valuable advice and support given by Susannah Roddick and other members of our Advisory Group at the Ministry of Education. We also wish to acknowledge the contributions of Deborah Widdowson for her assistance with Chapter 7 and Kathleen Murrow for her expert assistance with the final editing of this report.

Abstract

This report presents a literature review and conceptual model summarising the influence of peer effects on learning outcomes. We describe our approach to the review and provide a theoretical account of the environments, mechanisms, and processes that mediate learning among peers. We then summarise the literature on compositional effects at each level of school organisation--group, class, and school--and suggest how these effects might implicate peers by making connections to our theoretical account of peer-mediated learning. Next we make linkages across the different levels of inquiry in order to develop a conceptual model of peer influences on learning. We propose a multi-layered model with effects propagating from school-level influences to class-level influences to group-level influences to ambient and configured environments for learning among peers. We propose that the bulk of the effects are indirect. Hence, peer effects 'look' smaller the further we move away from the instructional coalface because they are mediated by intervening layers. We note that there may also be reciprocal effects whereby peers influence teachers and school organisation and management, though we are not sure of the magnitude of these effects. In concert with the three layers of influence, we argue that family resources have greater effects at upper layers and smaller effects at lower layers; conversely, curriculum and teaching resources have greater effects at lower layers and smaller effects at upper layers. Home and school supports for learning carry the lion's share of the weight in predicting student learning outcomes whereas peer effects, as currently constituted, carry much less weight. Finally we describe four instructional approaches that utilise peer resources to maximise learning. These models demonstrate additional ways of capitalising on peer effects beyond altering student composition.

Contents

Abstract ii
Contents iii
List of Tables v
List of Figures vi
Executive Summary vii
CHAPTER 1 1
  INTRODUCTION 1
    1.1 Approach 2
    1.2 Procedures for the Review 3
    1.3 Methodological Issues 4
    1.4 Organisation of the Report 11
CHAPTER 2 13
  PEER INFLUENCES ON LEARNING 13
    2.1 A Model of Peer Influences on Learning 13
    2.2 Peer Learning Environments 15
    2.3 Mechanisms and Processes by Which Peers Influence Learning 23
    2.4 Conclusion 28
    2.5 Recommendations for Further Research 29
CHAPTER 3 30
  GROUPING OF STUDENTS 30
    3.1 Grouping and Mixing Students by Ability 30
    3.2 Grouping and Mixing Students by Ethnicity and Gender 37
    3.3 Group Size 39
    3.4 Informal Grouping of Students 40
    3.5 Implicating Peer Effects on Learning 45
    3.6 Relevance to New Zealand 47
    3.7 Conclusion 49
    3.8 Recommendations for Further Research 50
CHAPTER 4 51
  CLASS CONFIGURATIONS 51
    4.1 Streaming 52
    4.2 Class Size 72
    4.3 Composite Classes 77
    4.4 Single-sex Classes 79
    4.5 Implicating Peer Effects on Learning 80
    4.6 Relevance to New Zealand 81
    4.7 Conclusion 82
    4.8 Recommendations for Further Research 83
CHAPTER 5 84
  SCHOOL COMPOSITION 84
    5.1 School Composition Broadly Defined 84
    5.2 Particular Dimensions of School Composition 96
    5.3 School Size 104
    5.4 Implicating Peer Effects on Learning 107
    5.5 Conclusion 109
    5.6 Recommendations for Further Research 110
CHAPTER 6 111
  MODELLING PEER EFFECTS 111
    6.1 Foundations 111
    6.2 Conceptual Model 113
    6.3 Accounting for Inconsistencies in Findings 117
    6.4 Conclusion 118

6.5 Recommendations for Further Research 119
Chapter 7 120
MAXIMISING PEER EFFECTS 120
7.1 Reciprocal Teaching 120
7.2 Collaborative Reasoning 122
7.3 Fostering a Community of Learners 124
7.4 Computer Supported Intentional Learning Environments (CSILE) 127
7.5 Conclusion 129
REFERENCES 130

List of Tables

Table 1.1
Equivalence Table Showing Relationship Between Effect Sizes and Success Rate Increase 7
Table 1.2
Effect Sizes for a Range of Typical School-based Factors 9
Table 1.3
Means, Standard Deviations, and Effect Sizes for the PATs Moderated by Gender 10
Table 3.1
Summary of 7 Studies of Within-class Grouping Reviewed by Slavin (1987) 31
Table 3.2
Summary of 15 Studies of Within-class Grouping Reviewed by Kulik and Kulik (1987) 32
Table 3.3
Summary of 11 Studies of Within-class Grouping Reviewed by Kulik and Kulik (1992) 33
Table 4.1
Summary of 23 Studies Reviewed by Kulik and Kulik (1984) 55
Table 4.2
Summary of 14 Primary School Studies Reviewed by Slavin (1987) 56
Table 4.3
Summary of 14 Joplin-Plan Studies Reviewed by Slavin (1987) 57
Table 4.4
Summary of 29 High School Studies Reviewed by Slavin (1990) 58
Table 4.5
Effect sizes for Studies Reviewed by Noland and Taylor (1986) 59
Table 4.6
Summary of 10 Studies Reviewed by Mosteller, Light, and Sachs (1996) 60
Table 4.7
Summary of Secondary Teachers' Qualifications in Low- and High-streamed Classes in Schools of Different Types (Oakes et al., 1990, p.66) 66
Table 4.8
Summary of Achievement Related Effect Sizes in 7 Meta-analyses 68
Table 4.9
Summary of Effect sizes Relating to Low, Middle and High-ability Streams in 7 Meta-analyses 69
Table 4.10
Stem and Leaf Display of Effect Sizes for Primary, Intermediate, and Secondary School 69
Table 4.11
Summary of Effect Sizes for Various Reductions of Class Size on Achievement and Attitudes 72
Table 4.12
A Comparison of Small and Regular Class Effect Sizes for Years 1 through 4 for White (W) and Minority (M) students in Project STAR 73
Table 4.13
Summary of Effect Sizes for Reading and Maths for Small and Regular Classes With and Without Teacher Aides for the SAT (Stanford Achievement Test) and BSF (Basic Skills First Test) 74
Table 5.1
Summary of British and United States Studies Reviewed 94

List of Figures

Figure 1.1 Distribution of Effect Sizes Across 50+ Million Students 8
Figure 2.1 Peer Learning Environments, Mechanisms, and Processes 14
Figure 6.1 Conceptual Model of Peer Influences on Learning 114

Executive Summary

This review is concerned with the influences of the composition of schools, classes, and small groups of students on students' learning outcomes and the contribution of peer effects to these influences. We examine compositions of students defined by ability, socio-economic status, gender and ethnicity in the general population. We also examine the effects of the size of schools, classes, and groups.

A conceptual model is developed to place the many peer influences into perspective. We note that peer effects can be considered as part of the surrounding ambient environment. We also identify a continuum of tutorially configured environments that mediate learning among peers ranging from learning that happens to occur in a social context (e.g., peer tutoring) to socially constructed learning (e.g. collaborative knowledge construction). Within these environments, we identify a range of cognitive and social mechanisms and processes that contribute to learning.

The results from our review, at all levels of school organisation, are that compositional effects measured in terms of effect sizes are small and show substantial variability, particularly at the school level. Theory and data converge on the notion that the effects of educational structures are small, indirect, and probabilistic in nature. That is, the effects of these structures (e.g., school mix, streaming class size) are mediated by an array of instructional and peer processes and the presence or otherwise of the structures serves to change the probability that certain cognitive and social mechanisms/processes occur (which then directly influence student learning).

Our review has revealed a pattern to the compositional effects. Not only are average effect sizes small--suggesting our indirect and probabilistic notions--they are relatively larger at the group level, somewhat smaller at the class level, and smaller still at the school level. Specifically, the small effect-sizes we have observed for within-class grouping range average about .25; for class-based influences, they average about .10; and for the more school-level influences, they average about .05 (although there is considerable variability at this level). If this pattern of effects is correct, then it lends credence to the notion of compositional effects operating through a series of nested hierarchical layers; effects are greatest at the 'coalface' where learning occurs and become smaller at the more distant layers. This is not to deny that there may be direct effects of school and class on individual learning; there may be, but we believe they are small and variable

In concert with the three layers of influence on learning, we argue that family resources have greater effects at upper layers and smaller effects at lower layers. Conversely, curriculum and teaching resources have greater effects at lower layers and smaller effects at upper layers. Home and school supports for learning carry the lion's share of the weight in predicting student outcomes and peer effects carry much less weight.

At the group level, research shows that there is a small but meaningful advantage of forming students into groups for instruction as opposed to using whole-class instruction. This seems to be true particularly when class sizes are large. Within groups, there is an advantage of homogenous over heterogeneous grouping though this result depends on the curriculum area and tasks. There is a risk that low-ability students' learning may suffer in homogeneous ability groups, not only from less instruction and less effective instruction but also from norms of behaviour that are not conducive to learning. These normative influences reflect a peer effect to the extent that peers contribute to a cycle of reciprocal teacher-student interactions that evolve over time.

In heterogeneous groups, depending on the composition of the group and other factors, students' ethnicity, and possibly gender, may also determine students' relative status and therefore their interaction and learning in the group. These peer effects stem directly from verbal interactions among students of higher and lower ability or among students whose characteristics are perceived as proxies for ability. We do not know, however, if the effects associated with ethnicity apply in New Zealand.

There seems to be some form of negative relationship between number of students in a group and learning outcomes though we are less convinced that this can be considered a 'peer effect'. This relationship may arise because teachers have to divide their time among larger numbers of students or because there is a lack of involvement of students in larger groups.

Peer interactions in informal groups inside and outside the classroom are associated with social and academic outcomes that may contribute to learning. In the case of informal talk and participation in extracurricular activities, we are reasonably confident that there are peer effects. However, in the case of playtime and lunchtime activities, we do not know if there are compositional effects, let alone peer effects, because the research has not adequately controlled for pre-existing personality and social differences between participants and non-participants, nor has it controlled for individual demographic characteristics.

At the class level, the effects of various class configurations average at best about .10. This estimate increases when teachers change their instruction to adapt more fully to the students in their classes. This change does not mean changing the pace of instruction or lowering the expectations of what the students can accomplish, but a dramatic change in the nature of the activities, a renewed vigour towards implementing appropriately challenging tasks, and implementing the many positive mechanisms that lead to enhanced student learning.

Whether a school streams or not, reduces class sizes, implements composite or single-level classes, or has coeducational or single-sex classes, appears less consequential than whether it attends to the nature and quality of instruction in the classroom, whatever the within-class variability in achievement. The learning environments within the classroom, and the mechanisms and processes of learning that they foster, are by far the more powerful. Good teaching can occur independently of the class configuration or homogeneity of the students within the class.

At the school level, we find great inconsistency in the effects of school composition in both the overseas and New Zealand literature. Overall, school composition accounts for a small amount of variance in student achievement, and where there are effects, they need sensitive designs to tease them out. We contend that many studies have not investigated, with sufficient detail, the degree of complexity involved in understanding school composition effects. Something important seems to be missing in the conception of how and when school composition might influence student achievement. We think a key problem is a lack of concern with theorising about what might be involved in creating school effects. This is because studies of school compositional effects on educational outcomes have used an input-output formulation and have, crucially, omitted to model how these influences act through the processes of schooling.

When examining schools at the 'extremes' of different student composition (e.g., public versus private, single-sex versus coeducational, home schooling versus public/private), it is difficult to establish to what extent advantages can be attributed to school composition. In these circumstances, school composition is inextricably interwoven with school resources, parental values, and students' background characteristics. There does, however, appear to be strong evidence for positive effects of Catholic private schools, at least in the United States, on learning outcomes, especially for low-ability students. These benefits may be the result of the communal organisational structure and inspirational ideology. Positive effects, especially for low-ability students, may also be attributable to the common core curriculum and higher academic focus in Catholic schools in the United States.

There is evidence that school size is a mediating factor in effects of school composition. Students appear to benefit from attending smaller high schools (around 800 students), as they appear to provide a more equitable learning environment. This is attributed to more personalised and intimate social relations. Small schools, because of their limited resources, are also more likely than larger schools to focus their resources on the provision of a core curriculum, which may also benefit students.

We have noted our findings of small effect sizes, particularly at the school and classroom levels, seem inconsistent with findings of descriptive studies (e.g., on school mix and streaming) that have not included measures of learning outcomes. One explanation for inconsistencies is that the findings from descriptive studies are correct—there are peer effects--it is just that the effects have little consequences for learning over and above those that can be accounted by individual characteristics of students. Another explanation is that there are peer effects but that outcome-based studies of compositional effects underestimate them because the operative level of peer effects is at a smaller level of aggregation than is typically studied. Another possibility is that there are peer effects but that outcome-based studies of compositional effects underestimate them because they do not do justice to the reciprocal relationships between students, teachers, and school organisation, and management. It could also be, as indicated earlier, that there are peer effects but that the outcome-based studies fail to model them in a theoretically appropriate way.

We present four instructional innovations that utilise peer resources and serve to maximise peer effects. All approaches draw from a theory of learning that emphasises the reflective and social nature of learning. All reflect a view of learning that is 'constructivist', where peers are seen as aiding in the construction of knowledge. There is evidence from these examples that these innovations promote collaborative processes, broadly conceived, and that these processes are linked to proximal indicators of learning such as increased interest and student engagement, increased production of higher-level cognitive processes, and enhanced learning outcomes. Policy directed at this level of analysis seems more likely to lead to enhanced learning outcomes than policy directed solely at the composition of students in schools and classrooms.

CHAPTER 1

INTRODUCTION

Research has identified family resources, school resources, peer effects, and individual ability as the main factors influencing learning, educational aspirations, and other outcomes of education (e.g., Zimmer & Toma, in press). The finding that peer effects influence learning outcomes has generated considerable interest as it suggests that social processes occurring in and around the school have important educational consequences. Policy decisions concerning various aspects of schooling seem to assume that peer effects exist. Assumptions about the importance of peer effects appear, at least in part, to underlie government policies on school enrolment and zoning, school policies on streaming of classes, parents' residential and school-choice decisions, male and female students' choice of subjects, and teachers' choice of grouping patterns within classes.

The finding that peer effects influence learning outcomes has come from studies showing that the composition of classes or instructional groups within classes has an independent effect on outcomes even after individual differences between students have been taken into account. There are also studies showing this effect for the composition of students within schools. These effects are variously termed 'compositional', 'contextual', or 'mix' effects, but most researchers interpret them as 'peer effects' (Willms, 1986). Studies show, for example, that measures of student ability, socio-economic status, ethnicity, or gender aggregated to the school, class, or group levels are related to achievement even when appropriate controls are introduced for students' individual scores on these measures. These 'peer effects' suggest that students who make up a school, or class, or group create a setting that facilitates or impedes learning above and beyond what would be expected on the basis of the individual characteristics of students.

There are various interpretations of these effects. Dreeben and Barr (1988) outline three explanations. One is a normative explanation that suggests peer effects arise because individual students internalise the norms of the educational setting to guide their learning and behaviour. Another is a comparative explanation that suggests peer effects occur because students use the educational setting as a reference group to make comparisons about their performance and develop academic self-perceptions. The third is an instructional explanation that suggests the effects occur because schools and teachers modify their instructional practices to take into account the characteristic of the student group. It is also possible that the ability or social mix of school influences the curriculum and the way students are streamed or guided into courses. Yet another explanation, not presented in Dreeben and Barr's analysis, is that peer effects arise from direct student-student interactions in schools and classrooms (Webb, 1991).

In this review, we examine the evidence on the effects of the composition of schools, classes, and small groups of students on students' learning outcomes in order to identify the contribution of peer effects. We examine compositions of students defined by ability, socio-economic status, gender, and ethnicity. We also take note of effects that are related to groups of students in particular curriculum areas (e.g., homogeneous groupings of students for reading or maths). However, we have not given special focus to particular ethnic groups in New Zealand (e.g., Maori and Pacific Island students) or to special populations (e.g., special needs or gifted students), since this was not our brief, except in so far as these groups are included in the groupings under study. In addition to examining the evidence for various groupings of students, we also examine the effects of the size of the school, class, or group. We do this under the assumption that, although size does not strictly relate to the composition of a student body, the size of a school, or class or group, may moderate compositional effects.

We provide a comprehensive review of the research literature relating to peer effects drawing from a range of disciplines including sociology, economics, education, and psychology. We survey both national and international literature and, in the case of the international literature, we analyse its relevance to schooling in New Zealand. Gaps and controversial issues in the empirical literature on peer effects are identified where appropriate. Finally, we develop a conceptual model of the influences of peer effects on learning outcomes and suggest ways teachers and other educators may capitalise on peer effects to maximise student learning.

1.1 Approach

The theoretical framework guiding our review is largely based on the work of Barr and Dreeben (1983) and Dreeben and Barr (1988). According to this framework, educational effects operate within a series of nested hierarchical layers of school organisation: community context and school, classroom, group, and individual students. This framework carries with it two implications. First, it specifies the levels of school organisation under which peer effects may be examined. The issue under investigation, then, is not how the composition at any single level of school organisation has an impact on instruction, but "how the composition of successive levels within the system shapes the arrangement of instructional settings" (Thrupp, 1999b, p. 34-35). Second, it assumes that the outcomes of each layer are best characterised as productive resources at lower levels of school organisation. Barr and Dreeben and others (see Bidwell & Kasarda, 1980) adopt a resource allocation perspective on the distribution of students to levels of school organisation, viewing student characteristics such as ability and socio-economic status as resources to be allocated to schools, classrooms, or small groups. So, for example, the distributional properties of schools are the outcomes of school-level composition and are best characterised as productive resources for classes; the distributional properties of classes are the outcomes of class-level composition and are best characterised as productive resources for instructional groups; and so on.

What this means for our analysis of peer effects is that each layer of school organisation has a different (but related) composition from those above or below it and that we need to examine how compositional effects occur at each layer of school organisation. According to Dreeben and Barr's (1988) formulation, a school "is engaged in a series of compositional transformations of its student population into grades, classes, and instructional groupings" (p. 132). It does this "to provide appropriate instruction [and curriculum] to a large and diverse clientele in aggregations of workable size and composition" (p. 132). Hence, for our analysis, we seek to build a cumulative compositional account for the entire school of how peers influence learning (cf. Thrupp, 1999b).

A corollary of the nested framework of Barr and Dreeben's (1983) model is that educational structures do not have direct influences on student learning. Rather, effects are mediated through successive layers of school organisation. We did not rule out the possibility that school, class, or group-level influences may have direct effects on learning, but we thought it unlikely. Ultimately, learning is a phenomenon that occurs in the individual, albeit mediated by peers and teachers. According to Barr and Dreeben (1991):

> *A social arrangement, in and of itself, does not lead directly to achievement or attitudinal outcomes; rather it is the activities and knowledge that students experience as part of instruction that bear directly on what they learn and how they feel about their learning (p. 895).*

Accordingly, in our review, we proceeded under the assumption that teachers bear the primary responsibility for shaping students' learning experiences, but that characteristics of students in a school, class, or group may set constraints on (or enable) the nature of the learning experiences that students encounter in these settings.

Hence, for the purposes of this review, we assumed that the relevant level of aggregation within which peer influences may be found was likely to be at the lowest level of school organisation. In other words, the framework suggested that the locus of peer effects was likely to be found in small groups of peers (pairs, cliques, or instructional groupings). This assumption is consistent with that of other writers on peer effects (cf., Harris, 1995; Moreland & Levine, 1992).

1.2 Procedures for the Review

The review was undertaken according to procedures designed to ensure that the information was relevant to the aims of the project and was complete in its discussion of both international and national research literature. Working from the objectives outlined in the original proposal, each of the core members of the team took responsibility for compiling lists of research relevant to their respective objective. This research was derived from sources cited in their own published research and from sources cited in seminal articles related to the objective. Some research was also obtained through informal collegial networks of researchers working in pertinent areas. All core team members read all seminal articles. In addition, computer-based searches were conducted using the major databases in the social sciences (these included *Australian Education Index, Current Contents, EconLit, ERIC, Expanded Academic, Index New Zealand, PsycLIT, Social Sciences Citation Index, and Sociofile*), supplemented by searches of local research reports. Searches were also made of a number of relevant electronic journals and associated web sites. Each team member then synthesised this material and took major responsibility for writing one of the Chapters 2 through 5 of this report.

In addition to assigning experts to each of the objectives, several features of the procedures provided further quality assurance. First, two research associates worked with all members of the team, across domains, assisting in aspects of search, synthesis, and writing. Second, core members were systematically involved in the development of the chapters just described by serving as critics on successive drafts being written by other members. In this process, conceptual linkages were forged across school, class, and group levels of analysis by reference to theoretical understandings of the mechanisms and processes by which peers influence learning. Once directions from the individual sections became apparent, emphasis was placed on the development of the conceptual model by the team as a whole. Third, team members were jointly involved in writing the additional chapters in this report, which provide a conceptual framework for examining the results of this review. This procedure ensured that chapters providing explanation or discussion concerning conceptual linkages were informed by input from all levels of the model. Fourth, all members of the team read the penultimate draft of the report and submitted written comments for discussion on the coherence and consistency of the arguments throughout the report. Fifth, within the restrictions imposed by a team project involving experts who enjoy academic freedom, the principal investigator took overall responsibility for promoting high integrity throughout the report, both in terms of the substance and the style. As a final issue of quality assurance, the report was reviewed and suggestions made for editing by a lay reader with skills in the presentation of technical reports for a wider audience.

This process was aided by a number of mechanisms that enhanced continuity, direction, and purpose. Regular weekly or twice-weekly meetings of the Auckland-based team were held, at which progress was reported and decisions made about subsequent directions. These meetings were subject to a prior-circulated agenda, which ensured that adequate time was dedicated to all objectives of the project. In the mid to latter stages of the project, a number of meetings, including a 'retreat' day, of the Auckland-based team (working with advice from the Bath members) were dedicated to theoretical discussions of the over-arching conceptual framework that would link all sections of the review, and the presentation of this framework in a visual model. Furthermore, all team members were involved in direct exchange of ideas via a listserv (edu-peer-lit@auckland.ac.nz) dedicated to the review. Monthly telephone conferences were held between the University of Auckland-based and University of Bath-based team members. One meeting was held with the research team responsible for the review on the effects of family and community resources on educational achievement, in order to share literature that was common to both reviews, and further contacts were made via email where relevant to the current work. Team members had formal meetings with representatives of the Ministry of Education in Wellington on two occasions, the first following production of a preliminary report (July, 1999), and the second after a more developed interim report (August, 1999). Although not able to be attended by the Bath-based members of the team, these meetings provided a mutual understanding of progress to date and directions for further development. Throughout the project, a research associate was in daily contact with team members on issues of time management, task completion, and coordination of activities.

1.3 Methodological Issues

In this section, we describe the evidence we used to identify compositional and peer effects in the empirical literature. We also describe the purpose and rationale of effect sizes, the metric we most frequently used to make comparative judgements of the magnitudes of effects, and the place of statistical significance testing in our interpretation of results from the research literature.

1.3.1 Identifying compositional and peer effects

We sought to identify peer effects on learning outcomes first by examining the evidence for compositional effects in various groupings of students defined by ability, socio-economic status, gender, and ethnicity. Then we relied on theory and data to determine whether the compositional effects we observed implicated peers. Throughout our review, we defined learning outcomes as indices of academic achievement or proximal indices of academic achievement. Proximal indices refer to measures of constructs that are on the causal path to, and are closely connected with, academic achievement (e.g., self-efficacy, self-concept, and achievement motivation).

For the purposes of this review, we defined compositional effects as the effect of the aggregate characteristic of a student group (e.g., mean level of ability) on a pupil's learning outcomes over and above the effects on learning associated with that pupil's individual characteristics. Despite the interpretations of compositional effect frequently made by researchers, described in the introduction to this chapter, compositional and peer effects are not synonymous. Compositional effects may arise from measurement artefacts in study design, differential school or classroom resources, differential school or classroom climates, and differential teacher practices, as well as peer effects. Differential resources, climates, or practices might be related to the composition of student groups but not strictly to peers. We define 'true' peer effects as the influences of student-to-student interactions and group dynamics on learning outcomes (where 'group' is used here in the sense of any aggregation of students, be it a pair, small group, class, or school). The term 'group dynamics' refers to the normative, comparative, and instructional explanations for peer effects described earlier. These dynamics may arise from cycles of reciprocal teacher and student influences that evolve over time and into which participants are socialised. Where peers are intimately involved in this socialisation process, we construe these dynamics as a 'peer effect'.

We sought evidence for compositional effects from experimental and correlational studies. The best evidence comes from experimental studies. In these studies, the performance of students organised into groups, classes, or schools is compared across different 'treatments'. These treatments might be 'homogeneously versus heterogeneously grouped students', 'streamed versus unstreamed classes', or 'single-sex versus coeducational schools'. If students in one treatment perform better than students in another, the inference that this effect is a compositional effect (i.e., goes beyond the individual characteristics of students) can be made to the extent that students in the two treatment conditions are equivalent on all relevant confounding variables. Students in the treatments are 'equated' preferably by randomly assigning students to the conditions. Failing randomisation, students are 'equated' by matching students on criteria related to the outcomes or, at least, by matching the groups, classes, or schools in which students are located.

The next best evidence, usually, comes from correlational studies employing regression techniques. In these studies, student performance on some outcome measure (the dependent variable) is regressed on a categorical variable (the independent variable) representing natural variation in the treatment of interest (e.g., streamed versus unstreamed classes). The extent to which a treatment effect can be interpreted as a compositional effect depends on how well variation on individual students' background characteristics (e.g., prior achievement, socio-economic status, gender, and ethnicity) is controlled. By including these background variables as independent variables in the regression equation, the unique contribution of the compositional variable over and above individual characteristics of students can be determined. The quality of the inference that a significant effect represents a compositional effect depends on the comprehensiveness of the background variables included in the model and the precision of measurement of these variables. The locus of the compositional effect (i.e., whether it represents differential teacher practices or a 'true' peer effect) depends also on the other variables included in the regression equation. Compositional variables are frequently student-level variables that have been aggregated to the group, class, or school level (e.g., mean ability of students in a school) and are subject to aggregation bias. In other words, the finding of a significant effect associated with a compositional variable may reflect other characteristics of the group, class, or school besides the collective of peers (unless, of course, these other characteristics have been taken into account in the equation).

A variant of the regression techniques used to identify compositional effects is a family of techniques known as multi-level modelling (MLM), the most common of which is hierarchical linear modelling (HLM) (Bryk & Raudenbush, 1992). These models explicitly take into account the nested structure of the data often used to identify compositional effects, where students are nested within classes and/or schools. They avoid the problems of misestimating the precision of estimates from conventional regression analyses, that occurr because students within the same grouping are more similar to each other than to students in different groupings. They do not, however, avoid the problem of aggregation bias. Within HLM/MLM, much as in conventional regression techniques, compositional effects are indicated when the aggregate variable makes a significant contribution in the between-class/school model predicting student learning outcomes, over and above the corresponding variable in the student-level model.

Where experimental or correlational studies provided good evidence of compositional effects, we relied on theory and data to determine whether they implicated peers. In some cases, researchers explicitly modelled the processes that underlie peer effects in their data (e.g., Webb, 1991). We found these studies convincing, particularly when the processes identified matched our theoretical understandings of how peer effects operate. In other cases, the nature of the controls in the experimental or correlational studies were sufficient to warrant making a claim that peer effects were involved, and we resorted to our theoretical understandings of mechanisms and processes mediating learning among peers to determine how peers might be implicated. In still other cases, neither theory nor data were sufficient to warrant making a claim about peers and we dismissed the findings as an artefact of uncontrolled variation (e.g., due to individual differences among students).

1.3.2 Effect sizes

One basis that provides comparative evidence of the various effects on learning outcomes is to calculate the effect sizes from various influences, including peer effects. Such a method begins by addressing the question "What are the 'typical' effects of schooling?" and then uses this typical effect as a benchmark for subsequent comparison of various influences on student learning. The problem is how to ascertain 'typical effects' given the myriad of effects on schools, different teachers, subjects, school administration systems, ages of students, and other moderators such as gender, prior ability, quality of instruction, and teaching styles.

The first requirement is a continuum on which the effects of schooling, including the typical effect, can be summarised, where zero means that there is no effect from introducing some teaching package, innovation, or configuration on schooling. A negative effect indicates that the innovation has a decreased effect on achievement, and a positive effect indicates that the innovation has an increased effect on achievement. For the present, the model is constrained to learning outcomes, but the continuum can be generalised to other outcomes of schooling.

The next requirement is to formulate an appropriate scale, and it is recommended that the scale is expressed in effect sizes. An effect size provides a common expression of the magnitude of study outcomes for all types of outcome variables, such as school achievement. An effect size is a measure of the direction and strength of the relationship between some factor (such as class size or streaming) and an outcome (such as achievement). It can be estimated from minimal information commonly reported in published research, and the most typical methods are as follows.

1. Comparing the means from two independent groups on an outcome measure (usually associated with a t-test or F-ratio) -- for example, the difference between the mean on an achievement test from a group of students in a streamed and an unstreamed class. The effect size, or z-score, is then calculated as d = (mean of the streamed – mean of the unstreamed class) / an estimate of the pooled standard deviation.
2. Conducting a study where a number of independent variables are correlated or regressed on some outcome. It is then possible to convert correlations to effect sizes by $d = 2r/(1-r^2)^{.5}$. For example, the correlation between self-concept and achievement is .2 (Hattie, 1992); therefore the effect size $= 2*.2\ (1-.04)^{.5} = .52$.
3. Evaluating the relationship between two dichotomies, such as when streamed and unstreamed students are compared on positive or negative outcomes. Usually a chi-square or variant is used, and this can be converted to an effect size by $d = 2(\chi^2 / (N-\chi^2))$. Thus, if $\chi^2 = 4.02$ with df = 1, based on a control group of 22 showing more success than a treatment group (n = 18), then the effect size = .67.

There are many other possible ways to estimate effect sizes and these have been outlined in the numerous sources available on meta-analysis (Cooper & Hedges, 1994; Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985).

The effect sizes can be averaged over a large number of studies to provide a best estimate of the typical effect, say, of streaming. A more powerful advantage of calculating effects is that they can be grouped by moderators to ask more detailed questions. For example, the effect sizes from the higher- versus lower-streamed students can be compared (e.g., d may be .4 for higher-streamed students and .2 for lower-streamed students, indicating twice the impact on the outcome for the former compared with the latter).

An effect size of 1.0 indicates an increase of one standard deviation, typically associated with advancing children's achievement by one year, improving the rate of learning by 50%, or a correlation between some variable (e.g., amount of homework) and achievement of approximately .50. When implementing a new programme, an effect size of 1.0 would mean that approximately 95% of outcomes positively enhance achievement, or average students receiving that treatment would exceed 84% of students not receiving that treatment. Cohen (1977) argued that an effect size of 1.0 would be regarded as large, blatantly obvious, grossly perceptible, and he provided examples such as the difference between the mean IQs of PhD graduates and high school students. The use of effect sizes highlights the importance of the magnitude of differences, which is contrary to the usual emphasis on statistical significance. Cohen (1990) has commented that "under the sway of the Fisherian scheme (or dependence on statistical significance), there has been little consciousness of how big things are; ...science is inevitably about magnitudes and the use of effect sizes makes a welcome force towards the cumulation of knowledge" (p. 1310).

Another method for interpreting effect sizes has been provided by (Rosentahl & Rubin, 1978, 1982). The binomial effect size display (BESD) addresses the question "What is the effect on the improvement of learning of an innovation?" It thus displays the change in improvement in learning rate attributable to the new learning innovation. Table 1.1 shows the equivalence between effect sizes and success rate increase. For example, if introducing computers into the classroom has an overall average effect size of .31 (which it does), this is associated with an improvement of learning from 42% to 58%, or a 16% improvement in the rate of learning. Compare this to the improvement rate of streaming, where there is an improvement of learning from 46% to 54%, or an eight-percent improvement. (It is sobering to consider that some of the major advances in medicine can also be placed on this effect size BESD continuum. For example, the use of aspirin to prevent heart attacks has an improvement rate of three percent.).

Table 1.1

Equivalence Table Showing Relationship Between Effect Sizes and Success Rate Increase

R	Effect Size	From	To	Success ratios
0.02	0.04	0.49	0.51	0.02
0.04	0.08	0.48	0.52	0.04
0.06	0.12	0.47	0.53	0.06
0.08	0.16	0.46	0.54	0.08
0.10	0.20	0.45	0.55	0.10
0.12	0.24	0.44	0.56	0.12
0.16	0.32	0.42	0.58	0.16
0.20	0.41	0.40	0.60	0.20
0.24	0.49	0.38	0.62	0.24
0.30	0.63	0.35	0.65	0.30
0.40	0.87	0.30	0.70	0.40
0.50	1.15	0.25	0.75	0.50
0.60	1.50	0.20	0.80	0.60
0.70	1.96	0.15	0.85	0.70
0.80	2.67	0.10	0.90	0.80
0.90	4.13	0.05	0.95	0.90
1.00	infinity	0.00	1.00	1.00

Hattie (1990, 1992) reported a synthesis of 337 meta-analyses, 200,000 effect sizes from 180,000 studies, representing approximately 50+ million students, and covering almost all methods of innovation. This synthesis can address the question "What is the typical effect of schooling?" The answer is derived from averaging the effects across the 357 meta-analyses and is .40 (with se = .05). Figure 1.1 depicts the distribution of all these effects across the 50+ million students.

Most innovations that are introduced in schools improve achievement by about .4 of a standard deviation. This is the benchmark figure and provides a 'standard' from which to judge effects, a comparison based on typical, real-world effects rather than based on the strongest cause possible, or with the weakest cause imaginable. At minimum, this continuum provides a method for measuring the effects of schooling.

The typical effect does not mean that merely placing a teacher in front of a class would lead to an improvement of .40 standard deviations. Some deliberate attempt to change, improve, plan, modify, or innovate is involved. The best available estimate as to the effects of schooling not based on innovations is from the National Assessment of Educational Progress (NAEP) data bank (Johnson & Zwick, 1990). NAEP surveyed what students in American schools knew and could do in the subject areas of reading, writing, civics, United States history, mathematics, and science. The students were sampled at ages nine, 13, and 17, and the testing has been repeated every two years. The average effect size across the six subject areas was .24 per year, which indicates that the effect of innovations is .16 (.40 - .24) standard deviations above and beyond the teacher effects. A further contention of many researchers is that maturation alone can account for much of the enhancement of learning. The effect of maturation is probably about one-third of the achievement effect (.10; see Cahen & Davis, 1987). Schooling does enhance learning above the influences of maturation.

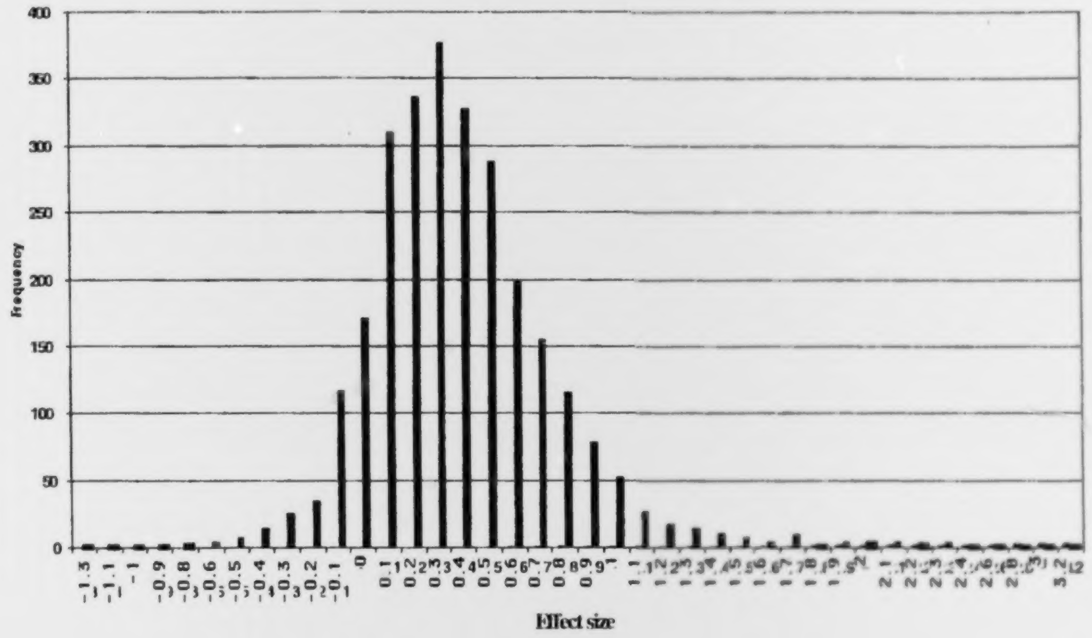


Figure 1.1 Distribution of Effect Sizes Across 50+ Million Students

Table 1.2 presents effect sizes for a range of typical school-based factors. It can be seen that the major effects relate to those introduced by some teachers. Of most relevance to the present study, the effects of peers are about 'average'. Of course, there can be critical moderators and mediators to these more gross overall summaries.

To provide an example of effect sizes, the means and standard deviations for the PAT Reading Comprehension, Reading Vocabulary, and Mathematics (Reid & Elley, 1991) tests are presented in Table 1.3. The effect sizes of the differences between girls and boys are also presented. In all three tests, it is noted in the manual that there are statistically significant differences between males and females, but, because of the inconvenience allowing for these would cause, norms are presented ignoring these gender differences. It is clear that the effects for reading are more marked in the junior years, but these differences disappear by the end of Form 4. For mathematics, the effects favour females in the junior years, but, by the intermediate years, come close to zero and slightly favour the males. These magnitudes would suggest that there should be different male and female norms for the Standards and that the genders can be combined for the Form years. The size of an effect is a more powerful indicator of the importance of any difference, it is easy to calculate, is more readily understood, and can assist in interpretation of the finding. These effect sizes are comparable to effects from other studies, as they are scale-free (that is, they do not depend on the scale of the original test).

Table 1.2

Effect Sizes for a Range of Typical School-based Factors

	No. of Effects	Effect Sizes
Reinforcement	139	1.13
Students' prior cognitive ability	896	1.04
Instructional quality	22	1.00
Instructional quantity	80	.84
Direct instruction	253	.82
Acceleration	162	.72
Home factors	728	.67
Remediation/feedback	146	.65
Students' disposition to learn	93	.61
Class environment	921	.56
Challenge of goals	2703	.52
Bilingual programmes	285	.51
Peer tutoring	125	.50
Mastery learning	104	.50
Teacher inservice education	3912	.49
Parent involvement	339	.46
Homework	110	.43
Questioning	134	.41
Peers	122	.38
Advance organisers	387	.37
Simulation and games	111	.34
Computer-assisted instruction	566	.31
Instructional media	4421	.30
Testing	1817	.30
Aims and policy of the school	542	.24
Affective attributes of students	355	.24
Calculators	231	.24
Physical attributes of students	905	.21
Learning hierarchies	24	.19
Ability grouping	3385	.18
Programmed instruction	220	.18
Audio-visual aids	6060	.16
Individualisation	630	.14
Finances/money	658	.12
Behavioural objectives	111	.12
Team teaching	41	.06
Physical attributes of the school	1850	-.05
Mass media	274	-.12
Retention	861	-.15

Table 1.3

Means, Standard Deviations, and Effect Sizes for the PATs Moderated by Gender

	Reading Comprehension				Reading Vocabulary				Mathematics			
	Form A		Form B		Form A		Form B		Form A		Form B	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Std 2 Mean	15.90	18.48	17.19	20.20	22.62	25.69	23.19	26.82	23.84	26.61	25.40	28.11
sd	8.23	9.26	7.86	7.73	10.59	10.17	10.95	10.48	11.26	10.10	11.42	10.31
effects	**0.30**		**0.39**		**0.30**		**0.34**		**0.26**		**0.25**	
Std 3 Mean	16.91	18.34	16.62	19.04	25.47	27.22	25.89	27.74	26.17	27.87	26.11	27.46
sd	8.39	8.06	9.03	8.45	11.74	11.13	12.54	11.67	10.18	9.34	9.90	8.86
effects	**0.17**		**0.28**		**0.15**		**0.15**		**0.17**		**0.14**	
Std 4 Mean	19.09	21.08	18.85	21.13	29.25	31.52	29.78	32.30	25.78	27.20	24.48	25.27
sd	8.76	7.93	9.52	8.70	13.30	12.01	13.84	12.71	9.98	9.59	10.13	9.59
effects	**0.24**		**0.25**		**0.18**		**0.19**		**0.15**		**0.08**	
F 1 Mean	21.41	22.72	22.25	23.75	34.37	35.82	36.70	36.94	28.52	29.48	26.62	27.60
sd	8.72	8.26	9.73	8.47	14.32	13.09	14.23	12.73	9.90	9.41	10.05	9.60
effects	**0.15**		**0.16**		**0.11**		**0.02**		**0.10**		**0.10**	
F 2 Mean	21.17	22.93	22.53	23.72	35.91	38.11	35.95	37.79	25.94	24.82	25.03	24.54
sd	8.93	8.63	8.75	8.11	13.20	12.91	13.36	12.41	10.30	9.26	10.49	9.28
effects	**0.20**		**0.14**		**0.17**		**0.14**		**–.11**		**–.05**	
F 3 Mean	24.21	24.96	24.54	25.24	37.17	39.54	37.13	37.36	23.18	22.97	23.66	23.70
sd	9.69	9.68	9.82	9.06	13.39	12.44	13.73	12.80	9.90	9.23	10.23	9.65
effects	**0.08**		**0.07**		**0.18**		**0.02**		**–.02**		**0.00**	
F 4 Mean	24.64	25.00	25.45	24.73	35.75	36.04	35.98	34.50	23.86	23.05	23.50	22.75
sd	10.00	9.53	9.09	8.65	11.90	12.42	11.96	12.05	10.27	8.99	9.80	8.62
effects	**0.04**		**–.08**		**0.02**		**–.12**		**–.08**		**–.08**	

It is possible to use effect sizes to make cost-benefit comparisons. For example, Levin, Glass, and Meister (1984) estimated the cost-benefits for four programmes to supplement mathematics or reading performance: reducing class size, increasing the length of the school day, computer-assisted instruction, and peer and adult tutoring. Although the costs were based on 1980 statistics, the relative information has probably not changed (although all details are provided in the paper to estimate costs based on 1999 New Zealand data). They estimated that the per-student costs for peer tutoring (based on costing ingredients and opportunity costs) was $212 for student participants (including tutors and tutees) and $827 for adult tutoring. The cost for introducing computers for a class of 32 students was $119. The cost for reducing class size from 35 to 20 students was $201 (and from 25 to 20 students was $94). For increasing instructional time in mathematics by 30 minutes per day (beyond 180 hours per year), the cost was $61. These are all additional costs only for mathematics or reading classes, and for adding these innovations and not replacing current teaching in the subject.

By comparing these incremental costs with the effect sizes from these interventions, Glass et al. (1984) were able to estimate cost-effectiveness ratios. These ratios were the effect size for each $100 of cost per pupil. The preferred alternative among the four interventions for increasing mathematics achievement was the peer tutoring model (at .46 per $100 per student invested) and the least preferred was increasing instructional time (gaining only .05 per $100 invested). For reading, the most effective was peer tutoring (at .22 per $100 invested) closely followed by computer assisted instruction (at .19 per $100 invested). Across both mathematics and reading the most effective was peer tutoring with an effect size of .34 per $100 invested. This was four times more cost-effective than reducing class size from 35 to 20 students. It is important to note that, while adult peer tutoring was associated with one of the largest average effect sizes (.67 for mathematics and .38 for reading) the costs were so substantial that it lead to one of the lowest cost-effectiveness ratios in mathematics (.08 per $100 invested) and the lowest in reading (.05 per $100 invested).

1.3.3 The place of statistical significance

The statistical significance test is a method by which a researcher can statistically evaluate the probability of obtaining an outcome against a theoretical reference distribution (often referred to as the null hypothesis). It is important to note that the results of a statistical significance test indicate nothing about an effect or provide information about magnitude. Rather, the statistical significance test addresses only whether or not the results one has observed may be explained by sampling variability. In other words, in performing a statistical significance test, a researcher is assessing whether or not the random sampling or random assignment procedure is a plausible explanation of the observed outcome. Conventionally, researchers declare outcomes with associated probabilities less than .05 as statistically significant. In stating this conclusion, researchers are inferring that it is rather improbable that the explanation of the observed outcome is the result of the random sampling or random assignment procedure (i.e. the random sampling or random assignment of participants within the experiment). Conclusions concerning the size or cause of an observed effect address analyses beyond the statistical significance test. Such considerations of effect size, in conjunction with the results of a significance test, paint a much clearer picture of research results (Abelson, 1995).

The use of statistical significance testing in the social sciences has repeatedly received much criticism in the literature. Many researchers contend that the statistical significance test is incompatible with the pursuit of knowledge in the social sciences. Cohen and others (Cohen, 1994, 1990; Gardner & Altman, 1986) have proposed alternatives to the significance test in an effort to improve the quality of studies (Carver, 1993; Gardner & Altman, 1986; Shaver, 1993).

A major problem with effect sizes is that they are independent of sample size. Hence, they may imply a sense of generalisability not warranted by the sample size. For example, an effect size of .5 based on a sample of 1,000 has a different strength of generalisability than an effect size of .5 based on 10 people. Hence, it seems desirable to have both statistical significance and effect sizes available to researchers. In meta-analyses, combining the various effects across many studies can offset the effects of small sample sizes, although there is still desirability in presenting sample sizes and effect sizes. The recent APA Task Force on Statistical Significance (Wilkinson, 1999) concluded that, when using effect sizes, "think of credibility, generalizability, and robustness. Are the effects credible, given the results of previous studies and theory? Do the features of the design and analysis (e.g., sample quality, similarity of the design to designs of previous studies, similarity of the effects to those in previous studies) suggest the results are generalizable? Are the designs and analytic methods robust enough to support strong conclusions?" (p. 602). These are the questions addressed in this report.

1.4 Organisation of the Report

Consistent with our view that peer effects may be best identified at lower levels of school organisation, we begin in Chapter 2 with a theoretical account of the environments, mechanisms, and processes that mediate learning among peers. This chapter lays the foundations for our review. We then summarise the literature on compositional effects at each level of school organisation and suggest how these effects might implicate peers by making connections to our theoretical account of learning among peers. Chapter 3 reviews the literature on group-level influences on learning. Chapter 4 reviews the literature on class-level influences. Chapter 5 reviews the literature on school-level influences. These three chapters summarise the empirical literature on compositional and peer effects and form the core of our review. In each chapter, we draw attention to the relevance of the findings to the New Zealand context and we make recommendations for future research. In Chapter 6, we identify issues that are common across the assorted literature and make linkages across the different levels of inquiry in order to develop a conceptual model of peer influences on learning. This model combines theory and data to present our best account of the trends emerging from our review. Finally, in Chapter 7, we note that there is more to peer effects than the composition of students and we describe four instructional innovations that point to additional ways of capitalising on peer effects to maximise learning.

CHAPTER 2

PEER INFLUENCES ON LEARNING

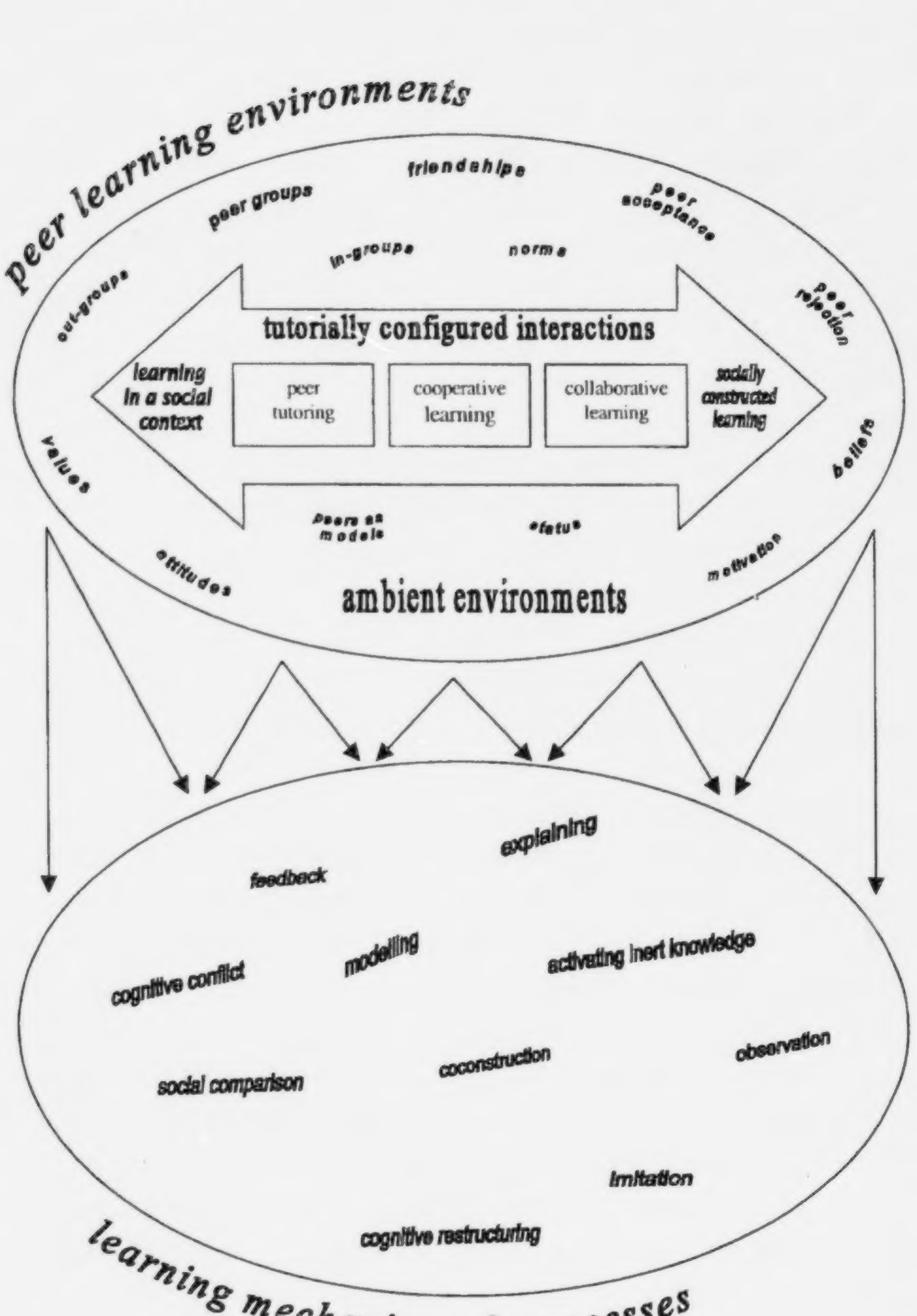
"I learned early on that my best teachers were my peers"
(a student, cited in Juel, 1996).

Current views of child development stress the significance of children in the lives of other children (Hartup, 1999). Similarly, those who follow a group socialisation theory of development emphasise the dominance of the peer group in children's development (e.g., Harris, 1995). The view taken in this chapter is that social influences are significant and that peer interactions in pairs and small groups are an important factor in learning. However, no assumption is made that the outcomes of learning are always positive. Peer influences may, for example, reinforce values not held by the school as a whole or by the larger society.

Peer group activity in learning settings has been seen to have effects on academic achievement, affective development, and social outcomes. But, despite the general acceptance of the influence of peer group learning, there is some uncertainty about how the dynamics and processes involved are related to learning (Cohen, 1994). Hartup (1999) has challenged the research community to develop theory and research that tell us more completely what happens in the course of social interaction between and among children or adolescents; "Whether children influence one another... is no longer in doubt. Critical issues, however, concern the manner in which subject and situational conditions interact with social contingencies in determining outcome" (p. 172). This chapter reviews literature in an effort to address this challenge. We begin by presenting a two-layer model of peer influences on learning that links peer learning environments with processes and mechanisms likely to be operating. We then elaborate on what we mean by peer learning environments and review the processes and mechanisms that are likely to mediate learning among peers.

2.1 A Model of Peer Influences on Learning

The two layers of our model are shown in Figure 2.1. The upper layer illustrates peer environments for learning. From the perspective of a socialisation theory of development, learning may result from interactions in environments which are broadly "tutorially configured" and from characteristics of the surrounding "ambient" environment (A third source of learning in a socialisation perspective comes from the development of personal systems. Since the focus of personal systems is less clearly linked with peers, it will not be discussed in this chapter). Tutorial systems are systems of learning and development that evolve in joint interactions. Configuration refers to the way in which the supportive and constructive nature of the interactions is structured to produce learning. (cf. McNaughton, 1995). The structure may include rules for participation or predictable sequences. The term ambient implies that features of the environment are available to the learner to observe and take implicit messages from but the learner does not engage directly as a participant and there is no formal attempt to instruct. In the natural environment children model their behaviour on others around them, take advice from a best-friend, adopt the values and beliefs of particular peer networks they join, and so on. Thus, observing others reading is an ambient influence, as is a shared belief within a peer group in the value of reading, or the social support offered by a friend in reading activities. The model represents our belief that the influences of ambient or naturally occurring environments are distinct from those that are configured. However, the ambient environment is portrayed as surrounding the configured learning environment to indicate that there are links between learning interactions in the configured environments and the processes operating in the ambient environment. Thus, learning in configured environments is always infused with the influences of the ambient environment. As Light et al. (1994) noted, "the effects of immediate face-to-face interaction between children working together on a task cannot be considered in isolation from the framework of social benefits and expectations that the participants bring to the situation" (p. 95).



BEST COPY AVAILABLE

Figure 2.1 Peer Learning Environments, Mechanisms, and Processes

As the model suggests, ambient environments are pervasive. Children spend significant time in the company of other children, so the potential for influence is great. Relationship dimensions such as friendships and peer groups may be associated with different influences on learning. Within a model of academic learning such as that proposed by Alton-Lee and Nuthall (1990; see also Nuthall & Alton-Lee, 1990), status among peers is seen as affecting access to resources – such as the availability of information or the extent of helping behaviour from other students. On the other hand, tutorially configured environments are shown in the model as occurring along a dimension or continuum (the "ends" represented by arrowheads) of interactive structure (McCarthey & McMahon, 1992). This dimension can be defined in terms of the extent to which the peer interactions are two-way. It can also be seen in terms of the extent to which knowledge is socially constructed as opposed to simply occurring within a social context. Thus, some configured peer learning environments, such as peer tutoring, are relatively low in two-way interaction and provide little opportunity for knowledge to be jointly constructed. On the other hand, configured peer learning environments such as collaborative learning have high levels of reciprocity between peers as they interact in the search for new, shared understanding. With respect to this dimension, Levine, Resnick, and Higgins (1993) make a similar distinction between interaction that stimulates thinking and interaction that constitutes thinking.

The lower layer of the model shows the mechanisms and processes by which peers influence learning. This includes both mechanisms and processes more commonly associated with tutorially configured environments and those associated with the ambient environment. Some examples of mechanisms and processes are noted. As suggested by the arrows between the layers in the model, there is no one-to-one mapping between peer learning environments and the individual mechanisms and processes; learning is seen as multiply determined. Recent work underscores the probabilistic relationship between the environments, mechanisms and processes (Cohen, 1994; Webb & Palincsar, 1996). There are certain conditions under which interaction can be structured to obtain maximal effectiveness. The structuring of interaction in certain ways affects mechanisms and processes. For example, variation in input characteristics such as task instructions, student preparation, and student and teacher roles foster different types of interaction and outcomes. Although the literature discusses mainly these three characteristics, in this report the focus is on compositional effects. In relation to interaction processes, perceived academic status differences are viewed as the most powerful, although differences in social status also affect interaction, notably participation (Cohen, 1982b; McAuliffe, Dembo, & Myron, 1994).

The format of this chapter follows the two layers of the model. First, the peer learning environments are outlined, beginning with the pervasive ambient influences on learning and followed by the configured learning contexts. In the second part of the chapter, we attempt to delineate the mechanisms and processes operating in peer learning environments more fully. Again, these are presented to reflect those most commonly operating in ambient environments and those more likely to occur in configured interactions with peers, although there is considerable overlap.

2.2 Peer Learning Environments

Within the social contexts for learning, research will be presented that discusses the effects that peers can have on learning outcomes or on close proximal indicators of such outcomes. In this section, mechanisms and processes are identified where appropriate, but section 2.3 has a more extensive discussion of these.

2.2.1 Ambient contexts for learning

Schools are places where students gather in self-selected or naturally occurring groups or associations that provide contexts for learning. Some of these groups are exclusive and stable best-friend pairs, whereas others are cliques and groups such as the 'in-group' or the 'out-group', which are much larger and more fluid in composition. These groupings both reflect and influence the behaviours, interactions, values, attitudes, and beliefs of students. The ambient environment is a more likely source of 'negative' learning effects such as low value for traditional academic goals, or higher desire to engage in counter-culture activities such as substance abuse. As seen in the model, every tutorially configured interaction takes place within an ambient environment that is infused with the influence of these associations. Thus, an understanding of the effects of these associations must be part of our understanding of peer influences on learning.

Friendships

Children's friendships represent one such association and are a prominent feature of the social landscape of a school. Friendships are based on similarities between peers, both in personal attributes (such as attitudes, values, activities, and personality) and in factors related to school organisation (such as gender, age, ethnicity, and socio-economic status). Motivation and academic performance at school are affected by friendships among peers (see Bukowski, Newcomb, & Hartup, 1998) as are school engagement, attitudes to school, and dropout potential (Wentzel & Caldwell, 1998). Children who have friends perform better at school than those who do not have friends (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Dishon, 1990; Frentz, Cresham, & Elliots, 1991; Krappman, 1985; Wentzel & Asher, 1995). Although many of these research findings are based on cross-sectional evidence, there is also longitudinal evidence that poor early friendship relations predict decreases in academic achievement, whereas making new friends in the classroom is linked to gains in school performance (Ladd, 1990). Friendship is also associated with a host of intrapersonal and interpersonal factors that may be considered proximal indicators of school performance – such as self-esteem, perspective-taking, communication skills, coping with stress, and so on (Bukowski et al., 1998). A New Zealand study (Patrick & Townsend, 1995) showed that perceived social competence with peers is an important precursor to the emergence of academic intrinsic motivation in Year 1 students.

Explanations of these effects have usually centred on the nature of friendship relations. One-to-one friendship is an egalitarian and supportive structure that allows intimate sharing and reciprocal disclosure between two people who have mutual liking and shared interests or activities (Aboud & Mendelson, 1996). In such contexts, friends provide new perspectives from which students "discover their own power to co-construct ideas and receive validation" (Seiffge-Krenke, 1993, p.76). Some support for this is found in Newcomb & Bagwell's (1995) meta-analysis of the behavioural and affective characteristics of friends and non-friends. This meta-analysis contrasted a variety of friendship status classifications found in the research studies, but typically the "friends" classification referred to studies involving reciprocal or mutual relationships while the "non-friends" classification referred to relationships involving acquaintances (classmates who are not selected as friends, but not disliked either), disliked peers, and strangers (children who do not know one another). From a search of three computerised databases (*Psychological Abstracts, ERIC, Dissertation Abstracts*), 82 studies were found which provided observational or non-observational (interviews, questionnaires, etc) data about the differences between eight types of friendship (e.g., reciprocal friend versus acquaintance; reciprocal friend versus disliked peer; unilateral-undefined friend versus stranger). A total of 524 dependent variables were coded in four 'broadband' dimensions of friendship relations: positive engagement, conflict management, task activity, and relationship properties. Friends were found to provide greater positive engagement (ie, social contact, conversation and discussion, cooperating, sharing, giving help, expressing affection), more frequent conflict resolution (ie, termination strategies, conciliation), and more effective task performance (ie, interactions related to the accomplishment of a task, resource utilisation, productivity) than non-friends. Similar notions have been expressed by Hartup (1998), who argued that friendship environments provide more effective learning through enhanced cooperation and greater understanding of the other's needs and abilities in problem-solving activities.

Peer groups

More research on peer influences on learning has been undertaken within the context of the larger peer group. Much of this research has focused on the effects of peers in influencing their schoolmates to engage in negative behaviours related to substance abuse and delinquency, particularly at the high school level. Less research has attempted to systematically document the way in which peers serve as a source of positive behaviour and attitudes in schools – such as those associated with valuing education, high achievement, and resistance to substance abuse.

Such research, however, does indicate that different groupings of peers may influence behaviour in different ways. For example, Urberg (1999) suggested that stable, best-friend relationships may be less important in influencing behaviours that are addictive or highly reinforcing (such as drug use, which is established in just a few trials), but much more important in influencing academic achievement, where there is often a long delay between the behaviour and subsequent reinforcement. An example from the primary school level is Kinderman's (1993) longitudinal study of the motivation to do well in school in cliques of children in fourth- and fifth-grade classrooms (Years 5 & 6). Children within cliques held similar attitudes towards school, but these attitudes were positive in some cliques and negative in others. As children changed cliques during the year (changing peer networks is common among younger children), their attitudes tended to change to more closely match those of their new group. Another differential effect of grouping is seen in the research of Bukowski, Hoza, and Boivin (1993), who found that lack of a best friend results in 'emotional loneliness', whereas lack of peer group acceptance results in 'social isolation.'

It is important to restate that peer influences can have negative outcomes. For example, high-achieving and low-achieving students may form different friendship networks that do not pursue the same goals, even within the same classroom. Wentzel (1989) and Finders (1997) found that high-achieving students have academic learning and social responsibility goals similar to those of the school. However, lower-achieving students adopt social attainment and group-allegiance goals that may be non-conformist, resist authority, and work against engagement with academic materials and tasks. Covington (1992; Covington & Beery, 1976) has argued that, in an effort to protect their sense of self-worth, low-achieving students may focus their classroom effort on self-handicapping behaviours that make them appear 'smart' to their peers but which do little to assist their learning, such as raising their hand to answer a question when they do not know the answer, procrastination in approaching a task, or deliberately not studying for a test. Observing their peers experience success as a result of their effort may heighten their own feelings of inadequacy and lead to the formation of peer groups that mutually reinforce less engagement with school tasks (Covington, 1987). With time, such children may develop peer networks whose members reinforce the attainment of social goals not valued by the school, such as those found in adolescent delinquency behaviours. Achieving such goals results in an enhancement of their reputation within their peer group community (Carroll, Durkin, Hattie, & Houghton, 1997; Emler & Reicher, 1995). Although the research on reputation enhancement is concerned with the more extreme forms of anti-social behaviour, there are parallels with the 'counter-culture' peer groups frequently found in schools. Such peer group effects may be more likely in settings that are more heterogeneous with regard to ability, socio-economic status, and cultural values, and where the school or class fosters a competitive academic environment.

These findings indicate that student attitudes, beliefs, and behaviours are influenced by natural peer contexts, but there is need for research which helps us understand the effects of peer influence as students move across different peer relationships (e.g., best friend, acquaintance, class clowns), different school settings (e.g., formal instruction, play-time), and different goal domains (e.g., academic, social, sporting).

In attempting to explain these social influences, broad mechanisms such as group socialisation (Harris, 1995; Wentzel, 1999) and internalisation (Hartup, 1999) are often invoked. Although the mechanisms and processes of peer influence will be discussed more fully later in this chapter, it should be noted that these concepts are not necessarily specific to school settings. Jencks and Mayer (1990) used similar concepts in their examination of the academic opportunities for children who grow up in poor neighbourhoods. They suggested that 'bad' neighbourhood peers influence other peers in a negative fashion by a 'contagion effect'. When applied to school, this contagion peer influence suggests that the existence of a cohort of students with poor academic achievement may create school norms for achievement that discourage other students from excelling. Alternatively, 'good' peers who have high achievement and who are well behaved may serve as positive role models to improve the outcomes for other students. Where such students are prevalent, as in higher socio-economic schools, peer networks might offer academic benefits to students who join such networks (Alexander, McDill, Fennessey, & D'Amico, 1979). Jencks and Mayer (1990) also argued that students engage in social comparison with those around them. In the school context this creates beliefs about what is expected of them as students. In schools that foster performance goals, such as doing well on national examinations and surpassing others, over mastery goals, such as improving competence and meeting self-set standards, social comparison may result in lower-ability students feeling worse off and performing more poorly.

Understanding the processes of peer influence is difficult, in part because specific mechanisms do not operate in isolation from other social or classroom influences. In the Kinderman (1993) study reported above, a change in person-to-group similarity may not simply be a function of peer socialisation influences of the existing members. Change may occur as a function of pre-existing conditions that create similarity, such as children being able to select their friends based on shared interests and values. Similarly, change may also result from influences outside the peer group. For example, some evidence suggests that teachers engage in differential behaviour towards students who are motivationally 'rich' or 'poor' and that this influences changes in students' behaviour and attitudes across the school year (Skinner & Belmont, 1993). To the extent that different types of students co-exist in natural groups, the change may be mistakenly attributed to a peer influence. For these reasons, some researchers have adopted different kinds of research strategies in order to unpack the mechanisms of peer influence.

One research strategy has involved the use of techniques that apply forms of causal modelling to examining peer influences. For example, Guay, Boivin, and Hodges (1999) used structural equation modelling to examine a hypothesised mediational pathway between friendship and achievement. These authors argued that perceived loneliness (an outcome of peer rejection) and perceived academic competence play a mediational role between peer relationships and change in academic achievement over the early school years. Using data from children (Years 3–5) in 10 schools in Canada, it was found that perceived loneliness influenced perceived academic competence, rather than the contrary, and that perceived competence, in turn, influenced changes in academic achievement over a three-year period. In brief, the more children were rejected by their peers, the greater their feelings of loneliness and social dissatisfaction; in turn, the more children perceived themselves as lonely, the lower their levels of perceived academic competence. Finally, lower levels of perceived competence resulted in a decrease in academic achievement. These results are consistent with, and link together, research which argues that friendships serve to increase feelings of competence in the classroom (Ladd, Kochenderfer, & Coleman, 1996) and that loneliness mediates the relationship between social status and self-worth (Boivin & Hymel, 1997; Sletta, Valas, Skaalvik, & Sobstad, 1996). Thus, the 'self system' processes of relatedness (seeing oneself as related to others in the school context) and competence (seeing oneself as being effective in school tasks) seem fundamental to an explanation of how friendships influence learning.

A different approach to identifying mechanisms of peer influence is to undertake a microanalysis of how changes in involvement in classroom work can be traced back to what is happening in interactions with others. For example, Sage and Kinderman (1999) carried out intensive observations of the social approval or disapproval expressed by 'social partners' when 'focal' children (within a single Year 6 classroom of 22 students) proceeded with the task in hand. Multiple observers coded over 12,000 behaviours for focal students and social partners across almost 500 observation sessions (each three minutes long) over 10 days. Peer group members and non-group members were found to provide different learning conditions during regular classroom interactions. Students who were highly motivated in achievement goals were more likely to be members of more highly motivated peer groups and to receive social approval from these peer group members following their active on-task behaviours. On the other hand, students who were less motivated were likely to be members of less-motivated peer groups and to receive social approval only from the teacher following on-task engagement. Students who were less motivated, and more frequently off-task, received more contingent disapproval from classmates who were not members of their peer group. The findings suggest that peer group approval for on-task behaviour, and non-member disapproval for off-task behaviour, can be a mechanism for positive change in highly motivated students and for maintenance in students with lower motivation. Longitudinal information or some other causal analysis is needed to confirm the proposed mechanism of social approval.

In summary, all learning in school takes place against a backdrop of ambient influences associated with peer group membership. These influences may operate differently in one-to-one egalitarian friendships than in wider peer network relationships. Researchers are agreed that student attitudes, beliefs, and behaviours are influenced by natural peer contexts. It is likely that much of the influence of peer effects is positive and in support of the goals and values of the school. However, much of the research on adolescent peer effects has focused on the negative outcomes of peer influence, such as delinquency behaviours and substance abuse. In spite of some recent innovations in research techniques applied to these issues, little research has attempted to isolate peer influences from other confounding influences, such as personal dispositions and influences from parents and teachers, and it is not possible to estimate the magnitude of peer influences stemming from the ambient environment. However, as evident from the research presented, investigations of peer effects are currently centred on the mediating social and cognitive processes of peer engagement with one another.

2.2.2 Tutorially configured contexts for learning

Current views of learning focus on its social and situated nature. The construction of knowledge occurs when the individual participates in situated social practices, and language is important in this process. Learning activities involving much communication and social interaction have been shown to be beneficial; collaboration and communication have been identified as key elements of effective instruction at all levels of education (Garcia, 1991). Where this collaboration and communication are explicitly fostered, we have delineated these to be tutorially configured environments. Within these environments, the emphasis is on interactions between peers. The teacher's role is typically to set up the group or peer context, give instruction and training, then let the groups or pairs operate relatively independently. There is, however, observation and monitoring of interactions and outcomes by the teacher and sometimes direct intervention to scaffold learning or to participate in the co-construction of norms (Webb & Palincsar, 1996). Configured interactions can differ in the explicitness of the support provided (Cazden, 1993; McNaughton, 1995), the extent to which the interaction is mutual, and whether the outcome is simply facilitated by the social setting or is socially constructed or co-constructed. These differences may be expressed in terms of a continuum, as shown in Figure 2.1. One end of the continuum represents learning environments that are structured settings in which the interactions are highly explicit and lacking in mutuality; peer tutoring is a tutorially configured environment that falls towards this end of the continuum. The other end of the continuum represents learning environments in which the outcomes are new understandings that have been co-constructed as an outcome of mutual, uncontrived interactions; collaboratively constructed learning is a tutorially configured environment that falls towards this end of the continuum. Cooperative learning is an example of a tutorially configured learning environment that lies between the two environments just mentioned. Although learning outcomes in cooperative contexts can be co-constructions, they need not be if interactions are not truly mutual, as happens when certain task responsibilities are 'owned' by certain group members, or when some members lead and others follow. Similarly, although cooperative learning may give students flexibility in the ways they engage and autonomy for their own participation, the use of cooperative learning in many classrooms is relatively contrived, with explicit goals and known resources which make co-constructed learning less likely. Much research has been conducted on these learning contexts, particularly peer tutoring and cooperative learning, and this research will be considered in the next section.

Peer tutoring

Tutorial interactions in peer tutoring, where one student has greater expertise, power, and control and assumes the role of tutor, are largely unidirectional. The direction of knowledge flow is from tutor to tutee (i.e., the student being tutored). In the process, the tutor may expand his/her own knowledge, but the tutee has little, if any, role in the construction. These interactions represent one extreme of a continuum in terms of how mutual the interaction is. Research evidence on peer tutoring shows significant achievement gains in reading, maths, and other areas (see Cohen, Kulik, & Kulik, 1982 for a review of 52 peer tutoring studies where the average effect size was .40). A note of caution is contained in the work of Shanahan and Barr (1995), who suggest that many of the studies where comparison groups are non-equivalent may be overestimating gains made by the children, particularly if the tutees were low achieving, because of regression to the mean. More recent research (Fuchs, Fuchs, Mathes, & Simmons, 1997a) shows that peer tutoring is effective, irrespective of student achievement and family income variables. Further, research has demonstrated that academic gains occur for both tutor and tutee (Simmons, Fuchs, Fuchs, Mathes, & Hodge, 1995).

Although research has largely considered cognitive aspects of peer tutoring, Shanahan (1999) reports work that has found peer tutoring to lead to more positive interpersonal relations or greater social acceptance among participants (Bowermaster, 1978; Eiserman, 1988; Jason, Erone, & Soucy, 1979; Larnen & Ehly, 1976). A collaborative peer tutoring programme, where each in a partnership takes turns to be tutor and tutee, was effective in improving self-concept and attitude to school (Roswal, Mims, Evans, & Smith, 1995). Studies have shown the importance of empathy and perceptions of caring on the part of tutors (Kaiden, 1994), although it is not clear how this affects the actions of the tutor or the responses and engagement of the tutee. Other studies have underscored the utility of training tutors (Bentz & Fuchs, 1996), particularly in how to offer and receive elaborated help (Fuchs et al., 1997b). In fact, Shanahan and Barr (1995), although referring to tutoring more generally, suggest that differences in outcomes among tutorial programmes may reflect differences in tutoring expertise in addition to the instructional programme characteristics.

According to Shanahan (1999), "we simply do not know why tutoring programs work, although there are many hypotheses and few empirical clues" (p. 229). Explanations proposed that are relevant to peer tutoring include greater individual involvement, improved attention, increased time on task, more appropriate individual pacing, more immediate and more relevant feedback, and greater opportunities for student identification with the tutor (Shanahan, 1999). Ginburgs-Block and Fantuzzo (1997) demonstrated higher rates of maths achievement, higher self-report levels of social acceptance, and higher observed teacher and student task-related behaviour in individuals who had experienced a reciprocal peer tutoring programme as compared to controls. A reciprocal or mutual peer tutoring model is somewhat different to the traditional peer tutoring, and King, Staffieri and Adelgais (1998) explain the gains in terms of the fact that role-switching engages students in a manner where emphasis is on questioning, explaining, monitoring, and regulation of learning.

Helping

Help seeking has only recently been considered a significant, adaptive, and productive strategy that aids learning. The ability of students to identify and utilise human resources in the classroom, particularly in the form of their peers, is a critical set of dispositions and skills (McCaslin & Good, 1996). According to Karabenick and Sharma (1994a, 1994b), it is a self-regulatory strategy. Those who come to school literate in classroom processes are more able to deal with classroom demands and may also be more able to seek help when they need it (Corno, 1989). Research on help-seeking in New Zealand intermediate schools has found that high-achieving Maori and Pakeha students report greater self-effort when in difficulty, but low-achieving students tend to seek help from the teacher; help-seeking directed to peers was both less likely and less preferred for classroom problems (Townsend, Manley, & Tuck, 1991). Students' home learning about such things as whether effort is more important than ability, and how responsibility is defined, influence students' attitudes to and understanding of using resources to help them achieve (McCaslin & Good, 1996).

The act of seeking explanation is likely to contribute to achievement, because it gives access to expertise and understandings that may be subsequently modelled, observed, and, later, imitated. Verbalising the request involves some rehearsal, and the act of verbalising or externalising thoughts in the form of a question makes those thoughts more accessible (King, 1990, 1992). Asking for and receiving explanations may lead to conceptual conflict with other students and the interaction may lead to cognitive restructuring. However, the relationship between seeking explanation and achievement is not clear. While some studies find a positive effect, some find no impact (Ross & Cousins, 1995).

Readiness to seek help has been shown to be related to academic self-efficacy (i.e., a belief in one's own capacity to reach a desired goal) in that those with low self-efficacy are more likely to associate asking for help with a perception by others that they lack ability, so they are less likely to engage in this behaviour. Also, help seeking is seen by some as a threat to self-worth and so is avoided (Ryan & Pintrich, 1997). This appears to be attenuated in classrooms where teachers believe that they should attend to students' social and emotional needs (Ryan, Gheen, & Midgley, 1998) and where classroom goal structures are seen to be emphasising self-improvement (Ryan & Pintrich, 1997).

Discussion groups and response groups

Recent literature (e.g. Mathes, Howard, Allen, & Fuchs, 1998) describes efforts to deal with diversity in primary age classrooms by employing what is termed 'peer mediation'. In the Mathes et al. study, a Peer Assisted Learning Strategies for First Grade Readers (PALS) programme was used. The reading-related outcomes were positively affected by participation in the programme. However, peer mediation was only one component of the programme, and other components included integrating phonological and alphabet skills into the decoding of words in context and providing extensive exposure to children's literature.

Other recent studies (e.g., Fuchs et al., 1997b; Fuchs, Fuchs, Kazdan, & Allen, 1999) show that explicit training of peers can lead to student interaction that incorporates more elaborated help giving. The Fuchs et al. (1997b) study resulted in tutors asking more participatory, procedural questions and providing more conceptual explanations. The achievement of those giving and receiving such elaborated help was higher. Giving elaborated help is thought to benefit the providers of the help by extending their own understanding and competence if they elaborate on that understanding. However, the literature is inconsistent in terms of the finding that elaborated help benefits the receiver. In the Fuchs et al. (1997b) study, students served in both roles, as tutor and tutee. In the later Fuchs et al. (1999) study, the gains in achievement for elaborated help, although substantial (mean effect size = .72), only held for older students (Year 5), while the younger students (Year 3) employed elaborated help less often and did better in the non-elaborated help situation.

There have been efforts to orchestrate discussion and response groups with students, on the Vygotskian (Vygotsky, 1978) premise that individual reasoning processes first emerge in interaction with others before becoming internalised by the individual. Drawing on the idea that discussions featuring reasoned argument among students have the potential for motivation and for improving reasoning, Chinn and Anderson (1998) explored a format called 'collaborative reasoning'. They outline two approaches to analysing the effects of argumentative discourse, the 'argument network' and the 'causal network', which they believe have implications as tools for improving instruction.

Likewise, students have been trained to provide appropriate questions and responses to pieces written by their peers so that the author may experience an audience and revise their work accordingly. Again, it is thought that the type of questions that peer respondents ask will be subsequently internalised by the writer and used to revise other pieces produced (Freedman, 1985).

Cooperative learning

Occupying a central location on the continuum of interactive structures are forms of cooperative learning where groups can be variously constituted. Cooperative learning tends to be known by a variety of terms which describe different methods; for example, *Group Investigation, Team-Assisted Individualisation, Jigsaw, Student Teams-Achievement Divisions, Learning Together, and Co-operative Integrated Reading and Composition* (Hertz-Lazarowitz, Kirkus, & Miller, 1995). The hallmarks, however, are that all participants are perceived as possible sources of expertise and that discourse is at times unidirectional and at other times multidirectional. The bulk of available research has concerned small-group interaction or cooperative learning where students work together in a group, participating in a collective task. Much research has favourably compared cooperative learning to other structures such as individual and competitive work (e.g., Johnson, Johnson, & Maruyama, 1983; Newmann & Thompson, 1987). Reviews of the research suggest that there are positive effects on achievement across school levels, ability levels, academic subjects, and types of skills when cooperative learning is incorporated as part of the teaching structure (Qin, Johnson, & Johnson, 1995; Stevens & Slavin, 1995). The origins of cooperative learning are in social psychology, and much research has focused on social outcomes. Cooperative learning has been shown to have positive effects on social outcomes, such as attitudes and helping behaviours towards classmates or workmates, the acceptance of ethnic minority members, and the acceptance of those who are 'different' (Johnson & Johnson, 1989). For example, research in New Zealand has shown the effectiveness of cooperative learning in increasing the peer acceptance of children with intellectual disabilities who have been mainstreamed in regular classrooms (Jacques, Wilton, & Townsend, 1998). Although these social aspects of the classroom are important, research on the application of cooperative learning in educational contexts has been primarily concerned with student achievement. Achievement is often found to be higher in cooperative learning contexts; however, there is variability in the research findings.

Variability in the outcomes of cooperative learning led Cohen (1994) to suggest that benefits can accrue only under certain conditions. Webb (1982, 1982b, 1984c, 1989b, 1991) has identified the most important predictors of peer interaction in such groups, the critical features of group interaction, and possible strategies for promoting effective small-group task-related interaction. Webb has shown, for example, that student experiences in group interaction depend on a combination of their own ability level and the ability composition of the group. In considering the findings from grouping studies, Lou, Abrami, Spence, Poulsen, Chambers, and d'Appollonia (1996) reviewed studies showing differential outcomes for high-, low-, and medium-ability students in homogeneous and heterogeneous groupings. They showed that lower-ability students have lower achievement in homogeneous than in heterogeneous groups, whereas medium- and high-ability students have higher achievement in homogeneous groups. The medium-ability students perform less well in heterogeneous than homogeneous groups (these findings are discussed in more detail in Chapter 3).

It is proposed that this effect could arise from different patterns of peer interaction. Webb suggests that in heterogeneous groups, high-ability and low-ability students formed an expert-novice relationship and medium-ability students tended to be left out of the interactions; whereas, in homogeneous groups, medium-ability students were active, often giving explanations to their peers. The work of Webb and colleagues places emphasis on giving and receiving explanations, particularly with regard to the level of elaboration of the help given and received. There is also an emphasis on requests for help and responses to requests for help as mechanisms for learning. In terms of giving help, variables that facilitated learning included giving content-related, elaborated explanations. Although giving help was beneficial for the student's learning, receiving help did not necessarily relate positively to learning. Rather, explanations had to be timely, correct, and sufficiently elaborated for the receiver to correct a misconception or lack of understanding. Webb, Troper, and Fall (1995) found that the strongest predictor of learning among those receiving help was the level of constructive problem-solving activity undertaken using concepts given in the explanations received.

Further, there is a complex relationship noted between the task environment and the interaction among peers (Cohen, 1994; Heap, 1986; Saunders, 1989). The type of interaction that is most effective varies with the nature of the task and the desired instructional outcome. For example, Cohen (1994) shows that the task instructions, level of student preparation, and teacher role that are suitable for supporting interaction in a routine learning task can constrain discussion in less structured tasks. Likewise, teachers' goals for, or academic performance expectations of, their students and teachers' responses to students (Bondy, 1990) may interact with other grouping variables.

According to Bossert (1988-89), three major mediating process explanations could account for the success of cooperative methods. These were: (i) 'reasoning strategies', where students are stimulated to engage in more higher-order thinking; (ii) 'constructive controversy', where students in heterogeneous groups have to accommodate the opinions of members and must engage in problem-solving and taking another's perspective; and (iii) 'cognitive processing', where cooperative work provides opportunities for students to rehearse material orally, to integrate it, and to explain how to approach a particular task. Others have theorised about the internalisation of social processes. The term 'appropriation' – extending and elaborating the 'internalisation' described by Vygotsky (1978) – has been employed by Rogoff (1993) to refer to the processes by which individuals actively transform their expertise through participation in learning activities with others, making the shared product their own.

Collaborative learning

Finally, at the opposite end of the interaction continuum to peer tutoring is collaborative learning. In true collaborative learning, knowledge is genuinely socially constructed between or among individuals. The knowledge is not held by one individual; it is sought and negotiated together so that the one collaborative outcome is greater than the sum of its parts. The discourse is bidirectional. While such learning is of great theoretical potential, by its very nature it is extremely difficult to examine, and there are few studies that investigate learning under these conditions (See Daiute & Dalton, 1993; McCarthey & McMahon, 1992). Research on collaborative writing by Daiute and Dalton (1993) indicated that some story writing knowledge was not simply transferred from more expert to less expert but, rather, co-constructed in the course of collaborative talk. Children demonstrated the use of a story element, for example, that neither had produced previously.

An excellent example of collaborative knowledge building is an effort to set up a learning environment that transferred the scientific model of knowledge building to a classroom setting. The scientific model involves public dissemination and discussion of findings that leads, iteratively, to the building of new knowledge, represented in public understanding. A computer supported intentional learning environment (CSILE) (Scardamalia & Bereiter, 1991, 1993-1994) was constructed to provide a medium so that knowledge as an object was visible and something that could be evaluated, examined for gaps, added to, changed and reformed. This medium took the form of a student generated common data base where everything entered was available to others. Student comment and discussion on the knowledge was a common activity. Students participating in such learning environments showed evidence of ability to ask questions concerned with explanation and to provide more depth in their explanations. Communal activity was significantly related to learning outcomes (for further discussion of CSILE, see Chapter 7).

In summary, in this section we have described configured environments structured to utilise peer interactions that potentially have the power to promote learning. The discussion of these environments has been presented to reflect the extent to which the interactions occurring within them are joint or mutual (the discussion of specific learning intervention programmes that have successfully utilised features of peer interaction is contained in Chapter 7). We have alluded to processes and mechanisms hypothesised or shown to facilitate learning within these environments and we now specify and categorise these in more detail.

2.3 Mechanisms and Processes by Which Peers Influence Learning

Within a social constructivist perspective, theories provide possible explanations of mechanisms by which social interaction facilitates development (Palincsar, 1998). Processes that operate between and among peers that directly or indirectly influence the individual's mechanisms that promote learning are related to structures within the learning context. As indicated previously, there is no one-to-one mapping between peer learning environments and mechanisms and processes. It is likely that each peer learning environment is influenced by mechanisms that are multiple, overlapping, and complex (Knight & Bohlmeyer, 1990). The following discussion of mechanisms and processes first considers those likely to be prominent in ambient learning environments, then those likely to operate in both ambient and tutorially configured environments, and, finally, those mechanisms and processes more associated with interactions in configured environments.

This section refers to 'mechanisms' and 'processes', and some consideration needs to be given to the use of these labels here. It is difficult to distinguish a process from a mechanism, as the two exist together and are closely related. Indeed, some researchers appear to use the terms synonymously. Hertz-Lazarowitz, et al. (1995) integrate the terms when they write of "cognitive process factors as underlying mechanisms" (p. 255). However, although it is often difficult to separate cognitive processes and mechanisms, there is some utility in the distinction. For example, cognitive restructuring is best thought of as a mechanism that can be triggered by the process of cognitive conflict or by the process of giving explanations. Thus, cognitive conflict is the context, which can be manipulated in a classroom, for the operation of the mechanism. In this sense, modelling is a process, even though it is often referred to as a mechanism. The mechanisms by which students learn from modelling are likely to be observation, imitation, and internalisation. Although research in social comparison has not demonstrated reliable gender differences, there may be a greater tendency for boys than girls to evaluate themselves relative to others (Huguet & Monteil, 1995; Joiner, Messer, Light & Littleton, 1998; Schwalbe & Staples, 1991).

Since the effects of the ambient environment are always present in tutorially configured environments, the processes and associated mechanisms that operate in one environment will also operate in the other. However, it seems likely that some mechanisms – such as peer feedback, observational learning from modelling, and social comparison – will operate more generally in the ambient environment. These mechanisms may have their effect not directly on achievement, but rather on proximal indicators of achievement, such as motivational constructs (e.g., self-regulated learning, beliefs about competence, and self-efficacious behaviour). On the other hand, cognitive conflict leading to cognitive restructuring, cognitive restructuring through giving explanations, and co-construction of ideas and activation of inert knowledge through production cues are more commonly associated with joint interactions that are tutorially configured, such as peer tutoring or cooperative learning groups. These mechanisms relate directly to enhanced academic outcomes. As mentioned previously, although the mechanisms are not restricted to either environment, the following discussion of mechanisms is structured so that mechanisms more likely to be associated with the ambient environment are discussed first, while the mechanisms discussed later are more likely to be associated with tutorially configured environments.

Social comparison

An aspect of modelling is social comparison, which occurs when students compare themselves with other students. This is an important means for learning about the appropriateness of behaviour, especially when the standards for such behaviour are ambiguous or unknown. However, social comparison also influences students' conceptions of ability and their evaluations of their own ability. Negative comparisons can have detrimental effects on low-performing students. Developmental research suggests that social comparison among peers is present as early as the first year of school in learning how to perform tasks, but is regularly used to evaluate personal competence by the time children reach about nine years of age (Ruble & Frey, 1991). Although research in social comparison has not demonstrated reliable gender differences, there may be greater societal pressure on boys than on girls to evaluate themselves relative to others. Opportunities for social comparison are greater in classrooms with a unidimensional organisation than a multidimensional organisation (Rosenholtz & Simpson, 1984). Unidimensional classrooms are characterised by an undifferentiated task structure, whole-class teaching or ability grouping, and public acknowledgement of achievement outcomes, all of which increase the salience of formal performance evaluations. In multidimensional classrooms, students are more likely to work on different tasks at the same time, often in interest-based (rather than ability-based) groups, making social comparison more difficult and offering less consistency to judgements of personal performance relative to their peers.

Affiliation through categorisation

In group socialisation theory, socialisation takes place in peer groups composed of individuals who characterise themselves in the same way. Generally, they share socially relevant characteristics such as age, gender, ethnicity, or, in adolescence particularly, abilities and interests. Behaviours and attitudes common to the majority in a group are accessible to the group as a whole.

Social norms

Individuals within a school share some common beliefs and values about the organisation, its purpose, its staff and students, and its teaching and learning practices. These core beliefs and values are constructed by the individuals and are reflected in their practices and behaviours (Blumenfeld, Hamilton, Bossert, Wessels, & Meece, 1983). Students must perceive, understand, and perhaps eventually internalise the conventional academic norms of classroom work (e.g., work hard, do your own work, complete homework assignments) and the conventional social norms (e.g., be quiet, help each other, be willing to work with all other children). To the degree that students perceive similarity between themselves and other students with regard to these values, adaptation to school is positive and the norms are reinforced. Some individuals, however, perceive themselves as different from the majority of students and do not endorse the beliefs of the larger group, thus leading to the development of a subculture with different norms. Subcultures are more likely to be formed when there is greater heterogeneity with regard to academic ability, ethnicity, and socio-economic status. In particular, students who doubt their capacity to achieve conventional academic or social success are more likely to lower their academic or social goals and set alternative goals designed to establish a non-conforming reputation – an extreme form of which is delinquency in adolescence (Carroll et al., 1997).

Socio-emotional support

School performance is affected by friendships among peers. In brief, friendships provide the social support (founded on reciprocity, commitment, and equality) necessary for good learning and development (Newcomb & Bagwell, 1995; Procidano, 1992). It is important to note that the positive outcomes associated with friendship appear to stem from aspects of intimacy (e.g., trust and loyalty), which may be achieved through one close friend, rather than from the size of one's peer friendship network (Townsend & Hansen, 1986; Townsend, McCracken, & Wilton, 1988).

Expectancy for success

A number of theories of academic motivation have distinguished between beliefs about being able to do the task (expectancy for success) and beliefs about the value of doing the task (task value), and have argued that it is the combination of the two that results in motivated behaviour for learning. For example, Eccles and her colleagues (Eccles, 1983; Wigfield & Eccles, 1994) have argued that the two most important first-level predictors of achievement behaviour are expectancy and task value. The expectancy construct is captured by the question, "Am I able to do this task?" As acknowledged by Pallas et al. (1994), expectancy for success has a strong influence on achievement. This relationship has been investigated in a series of large scale, correlational field studies involving cross-sectional and longitudinal methodology (Eccles, 1983; Wigfield & Eccles, 1992, 1994). Expectancies and perceptions of competence were the strongest predictors of achievement in mathematics and English (even better predictors than previous grades). Other researchers have linked expectancies and perceptions of ability to students' self-reported cognitive engagement – such as elaboration (paraphrasing, summarising) and metacognitive strategies (planning, checking, monitoring), which are involved in higher levels of learning and understanding (Pintrich & De Groot, 1990; Pintrich & Garcia, 1991; Pintrich & Schrauben, 1992). In the Eccles (1983) model, expectancy beliefs are a result of beliefs about personal competence related to the task and perceptions of task difficulty. These beliefs and perceptions are shaped by the child's experiences in the social world, including their interactions with school peers.

Task value

The second major predictor of achievement behaviour in the Eccles model is task value. The task value construct is captured by the question, "Why should I do this task?" This is not a simple question to answer, because a number of independent components of value are involved. First, there is an attainment value ("How important is it for me to be good at this task?"). There is also an intrinsic interest value ("How much do I enjoy doing this task?"). A third component is the extrinsic utility value ("How useful is this task going to be for me?"). Finally, these positive values must be weighed against any costs associated with engaging in the task (e.g., financial costs, time and effort, emotional costs such as fear of failure, and loss of opportunity to engage in alternative valued activities). Research indicates that task values for school subjects decrease as students progress through school (Eccles & Midgley, 1989; Wigfield, 1994a, 1994b; Wigfield & Eccles, 1992), particularly across the transition from elementary (primary) school to junior high (intermediate) school (Eccles, Wigfield, Flanagan, & Miller, 1989; Wigfield & et al., 1991). One explanation for this decline is that older children are more likely to engage in social comparison with their peers and are more able to integrate social comparison information with their own beliefs (Blumenfeld, Pintrich, Meece, & Wessels, 1982; Ruble & Frey, 1991). Since relatively few students excel at school, most students may be encouraged to lower their perceptions of the value of school subjects in which they are not particularly successful. Little is known about ethnic differences in task value, while gender differences appear to vary across age groups and curriculum areas (Wigfield & Eccles, 1992).

As noted above, one 'cost' associated with a task is the loss of valued alternative activities. Put simply, engaging in a homework assignment in mathematics may be at the expense of another valued activity, such as social interaction with friends. Research in New Zealand has shown that students who are more satisfied with their social relationships have higher task values for mathematics and language (Hicks & Townsend, 1993). Furthermore, both task values and social satisfaction are higher in classrooms using cooperative learning structures which address social needs through high levels of peer interaction (Townsend & Hicks, 1997).

Studies in the United States suggest that task values and expectations for success are each positively correlated with academic outcomes such as grades and standardised test performance. However, when both task values and expectancy beliefs are used to predict achievement, as in multiple regression studies, only expectancy beliefs survive as a significant predictor. At the same time, task values are better predictors than expectancy beliefs of which courses students will enrol in (Eccles, 1983; Meece, Wiggield, & Eccles, 1990, 1992). In short, task values may be more important for what students choose to study, but, once a choice has been made, achievement is more dependent on expectancy beliefs.

Social facilitation

'Social facilitation' refers to changes in behaviour resulting from the mere presence of peers. Interestingly, it is more than one hundred years ago that one of the earliest experiments in social psychology addressed the question of whether the presence of others influences an individual's performance. Triplett (1898) showed that competitive cyclists enhanced their performance when other cyclists were present. Social facilitation typically does not involve new learning, rather involving enhanced performance of existing learning. For example, a student may work more quickly when being observed by peers. More recent explanations for the effects of social facilitation suggest that it reflects a need for social approval (Geen, 1991). In children with behavioural difficulties, performance (number of correct responses and time taken) on cognitive tasks was found to be significantly influenced by peer presence (Bevington & Wishart, 1999).

Social loafing

'Social loafing' is the opposite of social facilitation and occurs where the presence of peers leads to a decrease in performance. As group size increases, performance decreases. Explanations of this effect include diminished responsibility for the outcome, reduced effort to achieve perceived equity of workload, reduced evaluation anxiety, and less clear standards of comparison (Geen, 1991; Latane, Willianms, & Harkins, 1979). Sanna (1992) has linked social loafing with self-efficacy and evaluation. Students with high self-efficacy perform better in a group setting when they are evaluated, whereas students with low self-efficacy perform better when not evaluated. This effect, in combination with the frequent use of groups in schools, has resulted in an examination of task and reward structures in groups (see research by Slavin, 1983 on cooperative learning).

Feedback

B. F. Skinner's (Skinner, 1968, 1953) operant conditioning theory has occupied an influential role in understanding achievement behaviour in classrooms. Students display learning behaviours that have been shaped by reinforcement and punishment, and for which effective reinforcers and punishers are available in their environments. Although most research has been concerned with the teacher's use of reinforcement, the influence of peer social reinforcement (e.g., verbal feedback, social acceptance behaviours) is critical in tutorially configured contexts (e.g., peer tutoring), and also highly likely to occur in ambient contexts (e.g., in informal friendship groups). However, rather than seeing reinforcers as response strengtheners, recent thinking suggests that reinforcement serves to inform students about the desirability and accuracy of their behaviour (Bandura, 1986). Thus, the effects of reinforcement are mediated through student cognitive processes associated with motivation (e.g., attributions, expectations for success, goals, and social comparison). To illustrate, in spite of the widespread acceptance of praise as a positive force in learning, it has rather weak associations with achievement, which are moderated by age, socio-economic status, and ability. Praise is weakly positive in influencing achievement in lower-ability students and those from lower socio-economic backgrounds, and weakly negative in influencing higher-ability students and those of higher socio-economic status (Brophy, 1981). This weak relationship most likely stems from relatively high levels of non-contingent teacher-praise that fail to substantiate students' beliefs that they are learning and which raise self-efficacy for learning; but it may also stem from inadvertent strengthening of inappropriate beliefs associated with ability and self-efficacy. When peer praise is given for success at an easy task, the praise may convey low expectations of the student, which lead to beliefs in lower ability and a subsequent negative effect on motivation to achieve.

It should be noted that peer feedback can occur formally within structured group situations. Within classroom structures such as peer response groups for writing, the presence of peers can make the idea of audience salient and 'audience' feedback can promote perspective taking by the writer (e.g., Freedman, 1985). Perspective taking leads to greater disclosure of information, better communication (through which information is phrased more accurately), better understanding of problems from multiple perspectives, better clarification of misunderstandings, and more positive attitudes to information exchange (Chalmer & Townsend, 1990). Personalised feedback from another person that enables an individual to compare performance to a standard increases performance and is especially important in attitude and behaviour change.

Modelling and observational learning

Similar mechanisms are at work in peer modelling effects. Peer models convey information about the functional value of behaviours and serve to motivate other students. Modelling refers to behavioural, cognitive, or affective changes that result from observing other people (Bandura, 1989; Schunk, 1987; Schunk, Hanson, & Cox, 1987). Modelling effects are more likely to occur if the peer model is competent, credible, and enthusiastic. They are also more likely to occur if the model is perceived (by the learner) to be similar to the learner. Watching similar others succeed at a task helps learners increase their self-efficacy and helps them believe that they, too, can be successful. This sense of efficacy motivates students to work on the task, and is validated as students experience success. Some evidence suggests that having access to a number of peer models is useful, as it increases the likelihood of obtaining a sense of similarity with at least one of the models. The evidence suggests that coping models, who initially display the typical fears and deficiencies of the observer, are more effective than mastery models, who display faultless performance and high confidence from the outset (Schunk, 1987; Schunk et al., 1987).

Cognitive restructuring: cognitive conflict

Cognitive conflict may be created by the divergent viewpoints of peers. Neo-Piagetian views emphasise the importance of social processes in an individual's knowledge construction (e.g., Perret-Clermont, Perret, & Bell, 1991; Tudge & Rogoff, 1989). The resolution of dissonance created by conflicting views in peer interaction is seen to influence individual intrapersonal processes. Cognitive restructuring results from incorporating ideas that contradict current schema.

Cognitive restructuring: providing explanations

The act of providing explanations to others (perhaps in the context of justifying personal views or simply sharing expertise) leads to greater understanding on the part of the giver and to demonstrable cognitive gain (Schwartz, 1995). The giving of explanations involves the explainer in clarifying and reorganising material and this cognitive restructuring may help the explainer to understand the material better, develop new perspectives, and recognise and fill gaps in his or her understanding (Webb & Palincsar, 1996). Learners conceptualise and organise material differently when preparing to teach someone else than they do when learning themselves (Bargh & Schul, 1980). The receipt of explanations, too, may lead to conceptual conflict, and the interaction may lead to cognitive restructuring on the part of the receiver.

Rehearsal

There is also a suggestion that framing a question to ask for help involves verbalising thoughts and some rehearsal. Rehearsal also occurs in the process of giving explanations and leads to enhanced memory for material.

Co-construction

The process of co-construction allows an individual to construct knowledge structures at a higher level than when working alone (Light et al., 1994; Schwartz, 1995). The mechanisms by which this occurs may involve transfer of expertise. Where a peer is more expert on a task or problem than other group members, he or she may use processes such as 'scaffolding' to transfer their expertise to the less expert members. This may involve the expert calling attention to salient features of task, providing feedback as to progress, and motivating the less expert to stay on task (Wood, Bruner, & Ross, 1976). Constructing knowledge at a higher level than when working alone may also come about through the mechanism of 'distributed cognition', whereby the strategic, processing, and knowledge load is shared. The knowledge of both procedures and content that the group has is greater potentially than that of any individual. An allied mechanism is one of exchanging resources.

Internalisation

Internalisation is a mechanism by which social dialogue is used as a source of problem-solving for the individual. For example, according to (Bershon, 1992), as students exchange ideas while working together, they have opportunities to build and use a vocabulary that directs and controls their activities during problem-solving. Then children internalise this language as what Vygotsky calls 'internal speech', developing a vocabulary to draw on to direct their actions in tasks.

Activating inert knowledge

Collaboration with a peer can activate 'inert' or 'passive' knowledge in a learner (Bruffee, 1973; Daiute & Dalton, 1993). A feature of collaboration that may expose inert knowledge is through the use of production cues such as occur in conversation. Bereiter and Scardamalia (1987) observed that, during independent writing, the lack of external cueing affected the discourse production in young children. They state that during conversation we are exposed to production signals and these signals may prompt and guide, in this case, discourse production.

2.4 Conclusion

In conclusion, research substantiates the idea that there are peer effects originating in peer interactions and associations. Some of these peer effects are 'institutional' and 'social', to use the terms of Pallas et al., (1994), and are considered more likely to operate in the ambient environment; other effects occur mainly in tutorially configured learning interactions such as structured peer tutoring and cooperative learning.

The best explanations for peer effects in learning are varied and may be found across a range of research disciplines. Mechanisms that the research suggests may explain the influence of peers in a social context for learning have been listed and described. These include mechanisms deriving from socio-cognitive theory, such as cognitive conflict and the associated cognitive restructuring and giving of explanations. Mechanisms suggested by socio-cultural perspectives include the transfer of expertise from expert to novice through scaffolding and distributed cognition. Consideration has also been given to the mechanisms that may influence proximal indicators of achievement, such as expectancy for success and academic task value. Peer effects may be explained through differences in the ways that groups of students are treated by other peers, and by how they react to or interpret these interactions. Thus, attitudes and beliefs about achievement may stem from peer feedback, peer modelling, and social comparison with peers, the outcomes of which are mediated through peer group norms and the socio-emotional support afforded by friendship patterns. Some of these mechanisms are more likely to be influential in ambient environments, while others are more likely to be influential in configured environments.

It may be that the influence of tutorially configured environments in learning is greater in primary schools where there is more grouping or seating to maximise peer help and cooperative interactions and where structured interventions such as reciprocal peer tutoring are employed. Some of the peer-mediated instructional interventions have been designed specifically to help deal with diversity in classrooms, which makes them more appropriate for primary classrooms since classrooms at this level are more likely to be heterogeneous in child characteristics. For example, informal ability grouping or pairing for different curriculum tasks allows the range of ability to be narrowed. Results from studies of reciprocal peer tutoring (Fuchs, Fuchs, Hamlet, & Karns, 1998) suggest that homogeneous pairings work more cooperatively and produce better quality work. On the other hand, there is little evidence of extensive use of small groups at the secondary level, suggesting that mechanisms and processes that operate in the ambient environment are more likely to be dominant at this level. This suggestion is supported by developmental theory, which postulates increasing influence of the peer group as children enter adolescence.

2.5 Recommendations for Further Research

- There is a need to establish the relative magnitude of peer effects on learning across various learning environments as they might operate in New Zealand schools. To date, there are only limited data on reciprocal teaching and peer tutoring.
- Research has yet to establish the causal structures that lead to learning in the various environments. For example, in 1999 we still do not really know why peer tutoring works. A corollary of this is the need to investigate the relationship between mechanisms and peer learning environments.
- It would be valuable in New Zealand to conduct research on the relationship between cultural influences and the mechanisms and processes associated with the peer learning environments existing in schools.
- Given the differential use of within-class groupings of students in primary and secondary school, there is a need to research differential patterns of influence from peer learning environments across school levels.

CHAPTER 3

GROUPING OF STUDENTS

In this chapter, we examine the extent to which the grouping of students affects their learning outcomes. We give primary consideration to grouping and mixing students by ability within classes. However, we also consider research on grouping and mixing students by ethnicity and gender, as well as research on the effects of group size. We also consider informal groupings of students inside and outside the classroom but within the school grounds. Where effects are found, we attempt to determine the extent to which peer effects are implicated and how peer effects might operate.

According to data collected from surveys for the International Association for the Evaluation of Educational Achievement (IEA), New Zealand primary school teachers are more likely to use grouping, and to use it frequently, than are teachers in most other Western countries. For reading, 94% of teachers of Year 5 students in New Zealand reported dividing their classes into groups for instruction (Wagemaker, 1993). Of these, 89% reported forming homogeneous groups on the basis of ability. For mathematics, all teachers of Years 4 and 5 students reported dividing their classes into groups at least some of the time, and 46% reported that they always used groups (Chamberlain, 1997). A similar pattern of usage is evident for science (Chamberlain, 1997). In most cases, these data refer to students working in pairs or small groups under the direction of the teacher. No data are available on the use of cooperative small groups by New Zealand teachers.

Results from the most recent meta-analysis of grouping (Lou et al., 1996) show a slight advantage of grouping compared to no grouping in promoting student learning (mean effect size = .17). Moreover, this analysis shows that the effect of grouping depends on class size. In large classes (more than 35 students), the mean effect of grouping is .35; whereas in small classes (less than 26), the mean effect is .22 and in medium-sized classes (26–35), it is .06. Small-group instruction has greater benefit compared to traditional whole-class teaching (mean effect size = .24) than compared to individualised mastery learning (mean effect size = .15). Also, small groups using cooperative learning perform significantly better (mean effect size = .28) than other small groups (mean effect size = .15). Low-, medium-, and high-ability students all seem to benefit from being taught in small groups (mean effects sizes = .37, .19, and .28, respectively). However, interpretation of these findings needs to be tempered by large variability in the effect sizes (see Lou et al., 1996). We think this variation can be explained, to a large degree, by differences in the instruction or academic task characteristics occurring within the groups and to differences in students' participation. Accordingly, we give weight to these considerations in the following sections.

3.1 Grouping and Mixing Students by Ability

Within-class ability grouping is traditionally viewed as a response on the part of teachers to the diversity of students' instructional needs. The practice was originally adopted "as a means for providing instruction that was more individually appropriate than total class instruction" (Barr, 1988, p.112). In theory, placing students into homogeneous groups on the basis of ability allows teachers to adapt the materials, level, and pace of instruction to the needs of individual students and, assuming the groups are flexible, allows continual adjustment to reflect changes in students' knowledge and skills. More recently, educators have also come to realise the pedagogical benefits that may occur from placing students into heterogeneous groups on the basis of ability (Bossert, 1988-89).

Consistent with the view taken throughout this review, we regard grouping as "important for individual learning not because of its direct influence but because of its shaping force on instruction, which in turn affects learning" (Dreeben, 1984, p. 74). Accordingly, we have structured this section in terms of the degree to which research has taken into account the instructional and normative features of small-group work (normative features refer to the ways students participate in relation to expectations established by the teacher). We begin with research that has simply examined group characteristics in relation to outcomes; we then proceed to review research that examines in more depth how group characteristics are related to the instruction given to students and students' social participation.

3.1.1 Effectiveness

A large number of experimental studies have investigated the effectiveness of different ability group compositions within classes on students' learning outcomes. Five major meta-analyses have been conducted that summarise findings from these studies (Kulik & Kulik, 1987, 1991; Lou et al., 1996; Slavin, 1987, 1990). The meta-analyses by Slavin and by Kulik and Kulik summarise the studies comparing the effects of homogeneous ability grouping versus no grouping (where teachers used whole-class instruction). Lou et al.'s meta-analysis summarises the studies comparing grouping versus no grouping, and those comparing homogeneous versus heterogeneous grouping.

Slavin's (1987) meta-analysis reviewed seven studies (eight experiments) of within-class ability grouping in primary schools in the United States. These studies are summarised in Table 3.1. In the tradition of Slavin's version of meta-analysis, called *best-evidence synthesis*, these were studies that met Slavin's methodological requirements for inclusion. In these studies, classes were randomly assigned to the two study conditions (grouped, not grouped), or they were matched on criteria related to outcomes, or students were matched within equivalent classes (among other criteria). All studies involved the use of maths groups, though one of these (Jones, 1948) also involved groups for reading and spelling. All but one of these studies involved students in the senior primary school (Year 4 and above). Results for every study showed that students in classes that were grouped for maths instruction performed better than students in ungrouped classes (median effect size = .34). These effects did not vary between the experiments using randomised and matched designs. There was no consistent pattern of effects of grouping for students of different ability levels (high, average, and low). Students at every level gained more in classes that used ability grouping than did their counterparts in the ungrouped classes. The median effect size for students of low ability was .65, whereas effect sizes for students of medium and high ability were .27 and .41, respectively.

Table 3.1

Summary of 7 Studies of Within-class Grouping Reviewed by Slavin (1987)

Study	Year levels	Duration (months)	Effect size			
			Overall	High	Medium	Low
Slavin & Karweit (1985)	4–6	4	.27	.41	.21	.29
Slavin & Karweit (1985)	5–7	5	.32	.13	.30	.65
Dewar (1963)	7	8	.55	.55	.43	.67
Smith (1960)	3–6	5	.41	.28	.25	.69
Wallen & Vowles (1960)	7	4	.07			
Spence (1958)	5–7	8	.44			
Jones (1948)	5	8	.26	48	.27	.37
Stern (1972)	4–5	4	.36			

The study by Jones (1948) is noteworthy because it involved reading and spelling groups as well as maths groups. According to Slavin, results of this study showed advantages of ability grouping for all subject areas. The effect sizes were .23 for reading, .43 for spelling, and .16 for maths. The results for each subject area could not be broken down by ability level of the student.

Slavin's later (1990) meta-analysis summarised the literature on the effects of 'ability grouping' in secondary schools, so the main focus was on effects of streaming (which is reviewed in the next chapter). However, two of the studies, both doctoral dissertations, were of within-class grouping in the secondary school. Campbell (1965) compared the use of three within-class groups for maths with whole class teaching in grade 7 (Year 8 students) in two Kansas City high schools. Harrah (1956) compared five types of within-class grouping in grades 7–9 (Years 8–10) in West Virginia.

Although Slavin does not report effect sizes for these studies, he reports that there were no significant differences in achievement between the various forms of classroom organisation in these studies. He notes that these results conflict with those of the studies of within-class grouping for maths in primary schools, which showed a positive effect of ability grouping for maths.

Kulik and Kulik (1987) conducted a meta-analysis of 19 studies of within-class grouping in the United States. Four of these studies are not included in the present review, as they examined programmes designed specifically for gifted and talented students. The results of the remaining 15 studies are summarised in Table 3.2. Seven of the studies were included in Slavin's (Slavin, 1987, 1990) meta-analyses. Most of the studies were carried out in primary schools. Some used intact groups for experimental and comparison conditions and did not control well for teacher and school effects. Kulik and Kulik noted that effects were significantly greater in studies where different teachers taught the grouped and ungrouped classes. Overall, the mean effect size in favour of within-class grouping was .17. Ten of the studies used within-class grouping for maths and these showed a mean effect size of .15. The study by Shields (1927) used within-class grouping for reading and the study by (Cignetti, 1974) used within-class grouping for typewriting. Kulik and Kulik also included Jones' (1948) study of within-class grouping for maths, reading, and spelling, as did Slavin (1987), but they estimated a slightly different effect size. Six of the studies reported results by ability level of the students. The mean effect sizes were .29 for high-ability students, .17 for medium-ability students, and .21 for low-ability students (these differences in effect size were not statistically significant).

Table 3.2

Summary of 15 Studies of Within-class Grouping Reviewed by Kulik and Kulik (1987)

Study	Year levels	Duration (months)	Effect size			
			Overall	High	Medium	Low
Bierden (1970)	8	8	-.16			
Campbell (1965)	8	8	-.16	.23	-.37	-.32
Cignetti (1974)	8–9	2.25	.09	.27	.22	-.41
Dewar (1963)	7	5.75	.36	.29	.29	.43
Eddleman (1971)	6	2.25	-.16			
Harrah (1955/1956)	8–10	4	-.03			
Jones (1948)	5	8	.29	.24	.27	.37
Mortlock (1969)	12	8	-.22			
Monroe (1922)	3,6,8	8	-.08			
Putbrese (1971)	5	8	.10			
Shields (1927)	8	1.50	.82			
Slavin & Karweit (1984)	4–7	4	.43	.41	.38	.50
Smith (1960)	3–6	4	.41	.28	.25	.68
Spence (1958)	5–7	7.50	.60			
Wallen & Vowles (1960)	7	4	.27			

Kulik and Kulik (1992) followed up their 1987 work by conducting a further meta-analysis that included 11 studies of within-class grouping, using more stringent criteria for study inclusion and somewhat different procedures for calculating effect sizes. These studies are summarised in Table 3.3. All of the studies were from the earlier review. Eight of the studies were carried out in primary schools and three in secondary schools. All but one used standardised tests as outcome measures. Again, some used intact groups for experimental and comparison conditions and did not control well for teacher and school effects. Overall, the mean effect size in favour of within-class grouping was .25. Eight of the studies used within-class grouping for maths and these showed a mean effect size of

.21. Six of the studies reported results by ability level of the students. The mean effect size was .30 for high-ability students, .18 for medium-ability students, and .16 for low-ability students (again, these differences in effect size were not statistically significant).

Table 3.3

Summary of 11 Studies of Within-class Grouping Reviewed by Kulik and Kulik (1992)

Study	Year levels	Duration (months)	Effect size			
			Overall	High	Medium	Low
Campbell (1965)	8	8	-.18	.26	-.41	-.36
Cignetti (1974)	8–9	2.25	.09	.27	.22	-.41
Dewar (1963)	7	5.75	.48	.47	.50	.56
Eddleman (1971)*	6	2.25	-.09			
Jones (1948)	5	8	.21	.19	.23	.40
Putbrese (1971)	5	8	.16			
Shields (1927)	8	1.50	.82			
Slavin & Karweit (1984)	4–7	4	.43	.41	.38	.50
Smith (1960)	3–6	4	.22	.18	.15	.30
Spence (1958)	5–7	7.50	.60			
Wallen & Vowles (1960)	7	4	.06			

* This study compared homogeneous and heterogeneous grouping

The more recent meta-analysis by Lou et al. (1996) examined results from 51 studies (103 contrasts) on the effects of grouped versus ungrouped classes at the primary, secondary, and post-secondary levels, without regard to the composition of the groups (homogeneous or heterogeneous). Results from this comparison were reported in the introduction to this chapter. This meta-analysis included all studies previously reviewed by Kulik and Kulik (1987, 1991) and Slavin (1987, 1990) but included additional studies. The mean effect of homogeneous grouping compared to no grouping (.16) was similar to the mean effect of heterogeneous grouping versus no grouping (.19), though there was significant variability in each set of findings.

Lou et al. (1996) also summarised results of 12 studies (20 contrasts) comparing homogeneous versus heterogeneous grouping (effect sizes for individual studies were not reported). This comparison showed a slight advantage of homogeneous over heterogeneous grouping (mean effect size = .12). The advantages of homogeneous over heterogeneous grouping were greatest in reading (mean effect size = .36) – though this finding is based on only four effect sizes – and the effects were non-significant in maths and science (mean effect size = .00). Moreover, results showed that medium-ability students benefited the most in homogeneous groups when compared to their performance in heterogeneous groups (mean effect size = .51), whereas low-ability students faired worst (mean effect size = –.60) and high-ability students showed no meaningful difference (mean effect size = .09). The negative sign for low-ability students indicates that these students actually performed better in heterogeneous groups.

One correlational study showed that, although ability grouping can have positive effects, there are tradeoffs. In an intriguing study, Sorensen and Hallinan (1986) examined growth in reading achievement of 564 students (Years 5–8) in California based on 384 students in classes that used homogeneous within-class ability groups for reading and 180 students in classes that used whole-class instruction. Their results suggested that students in the small-group lessons learnt more of what was taught because the instruction was better adapted to students' abilities and/or they were more attentive. However, they found that ability grouping provided fewer *opportunities for learning* than whole-class instruction because students in grouped classes were given smaller amounts of instructional time (since it had to be divided among the groups). Sorensen and Hallinan's findings also suggested that grouping increased the inequality of learning outcomes. Students in high-ability groups were given more opportunities for learning than were students in low-ability groups. Results from this study need to be interpreted carefully as students in the grouped and ungrouped classes were not equated in terms of prior abilities and, as with the experimental studies, we do not know the exact nature of the instruction and students' participation.

A consistent message from studies of the effectiveness of grouping and mixing students within classes by ability is that instructional materials and the nature of instruction must be adapted for small-group learning. Simply placing students in small groups is not enough. For grouping to be maximally effective, particularly in the case of homogeneous groups, materials and teaching must be varied to accommodate the needs of students of different levels of ability. This may be especially important for low-ability students in homogeneous groups where the demands placed on students by the materials and level of instruction seem to be crucial to growth in achievement.

In summary, the available evidence shows that there is a small advantage in forming students into homogeneous ability groups as opposed to using whole-class instruction, at least in primary school. Slavin (1987) reported a median effect size of .34; Kulik and Kulik (1987) reported a mean effect size of .17; Kulik and Kulik (1992) reported a mean effect size of .25; and Lou et al. (1996) reported a mean effect size of .16. Lou et al. (1996) also reported a mean effect size of .19 for the advantage of using heterogenous groups over whole-class instruction. The differences in results for homogenous groups can be explained by differences in the samples of studies included in the meta-analyses and to differences in the way effect sizes were estimated. The latter is worrisome; it is very disturbing to find effect sizes for the same study differing between researchers (sometimes even between analyses by the same researchers) by a factor of two or more. Nevertheless, all meta-analyses show small positive effects of within-class grouping and they show that students at every level gain more in classes that use ability grouping than do their counterparts in ungrouped classes. The meta-analysis by Lou et al. goes further to show a slight advantage of homogeneous over heterogeneous grouping, though this result depends on curriculum area and probably task. The advantages of homogeneous grouping seem to be greatest in reading and less in maths and science. However, it should be noted that most of the studies of within-class grouping, whether homogeneous or heterogeneous, relate to grouping for maths. Fewer studies have investigated the effects of within-class grouping for reading, but what evidence there is suggests that the effects are quite positive.

3.1.2 Instruction and social participation

There has been much naturalistic research describing the instruction and social participation found in groups of different ability. Most of this research has focused on homogeneous ability groups for reading (for reviews, see Allington, 1983; Barr & Dreeben, 1991; Bloome & Green, 1984; Hiebert, 1983). This research comes closest to describing possible peer effects, but the studies have not included measures of students' learning. Many of the studies are case studies or ethnographies, and the researchers have relied on theory and logic to make inferences about the likely effects of instruction and social participation on learning. By and large, the research suggests that ability grouping results in differential learning experiences that seem to perpetuate, or even exacerbate, inequalities among students. Research shows, for example, that groups may be rigid and restrict mobility of students between groups, and that teachers provide less instruction and less effective instruction for children in lower-ability groups. Teachers with lower-ability groups may allocate less time to lessons, conduct lessons at a slower pace, and generally emphasise lower-level rather than higher-level tasks. Moreover, this research suggests that groups develop subcultures or norms of behaviour that differentially support learning; there is greater inattention among children in lower-ability groups, less of the talk is about the task at hand, and there are more call-outs and other interruptions to the lessons.

Looking firstly at the time teachers spend giving instructions to the groups, the studies generally show that teachers allocate less instructional time to low-ability groups than to high-ability groups and to allocate time to different tasks for the different groups. Hunter (1978), McDermott (1976), and Rist (1970) reported that teachers spend less time with low-ability groups (though some studies have not found differential time allocations, see Collins, 1986; Grant & Rothenberg, 1981;

Weinstein, 1976). Moreover, teachers with low-ability groups have been found to spend more time on decoding tasks that focus on individual words and parts of words and less time on tasks that relate to meaning (Allington, 1980a, 1980b, 1984; Alpert, 1974; DeStefano, Pepinsky, & Sanders, 1982; Duffy & Anderson, 1981; Gambrell, Wilson, & Gantt, 1981; Hart, 1982; McDermott, 1976; Stern & Shavelson, 1981). Teachers with low-ability groups also spend more time on oral reading than silent reading compared to high-ability groups (Allington, 1977, 1980a, 1980b, 1983, 1984). The difficulty level of assigned tasks also varies across ability groups. There is some evidence that low-ability readers are required to read material that is more difficult for them, relative to their reading level, than the material given to high-ability readers (Alpert, 1975; Berliner, 1981; Gambrell et al., 1981).

In addition, the pace of instruction tends to be slower in low-ability groups (Allington, 1984; Barr, 1982; Barr & Dreeben, 1983). Hence, children in low-ability groups read less than do children in high-ability groups. Allington found that teachers moved low-ability readers at the pace of only one segment of a story per lesson, whereas high-ability readers usually completed one complete story per lesson. This meant that the children in low-ability groups read only half, and sometimes as little as a third, as many words as the children in high-ability groups. In part, this may be because more of the reading with low-ability groups is oral rather than silent. Oral reading is somewhat slower than silent reading, so children read less in a given time if they read out loud.

When teacher-student interactions are studied, research shows that teachers are more likely to interrupt poor readers who make oral reading errors than they are good readers who make similar errors (Allington, 1980a, 1980b). Also, the feedback teachers give in response to the errors is more likely to be 'terminal feedback' – that is, telling the student the correct word – rather than 'sustaining feedback' – that is, prompting the reader to correct his or her own error (Anderson, Evertson, & Brophy, 1979; Hoffman et al., 1984). Moreover, when sustaining feedback is provided to children in low-ability groups, teachers are more likely to give a clue about the graphemic or phonemic characteristics of the target word rather than directing the children's attention to semantic and syntactic information (Allington, 1980a, 1980b). When discussing the material read, teachers ask relatively more simple, factual, recall questions of students in low-ability groups and relatively fewer questions that require reasoning (Seltzer, 1976). Once a question has been asked, students in the low-ability group may be given less time to answer (Good, 1981). There is also some indication that students in low-ability groups are given proportionally less praise for correct responses than are students in high-ability groups (Seltzer, 1976) (Alpert, 1974; Grant & Rothenberg, 1981; Martin & Evertson, 1980).

A large number of studies investigating the nature of students' social participation in groups have found that students in low-ability groups are less engaged than students in high-ability groups (Gambrell, 1984; Gambrell et al., 1981; Good & Beckerman, 1978; Haskins, Walden, & Ramey, 1983; Martin & Evertson, 1980; Metz, 1978). These effects have been found even after controlling for individual characteristics of students. In addition, less of the talk in low-ability groups tends to be about the task at hand (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981), and students in low-ability groups communicate less with their teachers (Rist, 1970). Grant & Rothenberg (1981) also found that teachers allow more interruptions when working with their low-ability groups than when working with high-ability groups. Similarly, Camp and Zimet (1975) found that teachers spend a larger proportion of time dealing with students' behaviour and attention problems in low-ability groups than in high-ability groups.

Despite the association between membership in low-ability groups and what appear to be less effective instructional practices, experts on grouping stress that grouping practices and instruction ought to be thought of as distinct but interacting factors (Barr & Dreeben, 1991; Hallinan, 1984). Grouping practices relate to the way students are assigned to groups for instruction. They include decisions about the basis for assigning students to groups: the number, size, and composition of groups; and the stability of group membership over time. Instruction relates to the way students are taught within groups: the methods of instruction, resources and materials, pacing, and means of evaluation. In Hallinan's words:

> *Grouping practices and mode of instruction may vary separately or jointly across tracks or group levels... Ultimately, it should be possible to determine which particular grouping practices are most effective at each ability level and how their effectiveness relates to the mode of instruction that distinguishes that level. Thus, research models should specify how grouping practices and modes of instruction interact to enhance or hinder learning (pp. 234–235).*

Dreeben (1984) noted that grouping and instruction are not only conceptually distinct, they are also empirically distinct. Citing evidence on pacing and content coverage from Barr and Dreeben's (1983) study of learning in small-group reading lessons, he noted that groups – even ones similar in composition – may receive "vastly different instructional experiences" (Dreeben, 1984, p. 74). He concluded by saying:

> *There is little point in praising or condemning grouping per se for its instructional and social consequences. It can be used well or ill. The consequences of grouping are at least as much a function of how groups are used as of their mere existence* (p. 83).

Nevertheless, more fine-grained analyses show that teacher's instructional strategies may interact with students' behaviour and processing to produce differential effects. This research suggests that there are reciprocal influences between students and teachers, and students and teachers are acculturated or socialised over time into group-specific norms of behaviour.

McDermott (1976) was among the first to notice that the instructional process is a collaborative one, where teachers and students build upon one another's verbal and non-verbal behaviour. "This collaborative process unconsciously creates a pattern of interaction that is either harmonious and directed at reading or disharmonious and filled with interruption" (Collins, 1986, p. 119). McDermott conducted an ethnographic study of teacher-student interaction in the life of one first-grade (Year 2) classroom in the United States. The major portion of the study was a detailed analysis of the moment-to-moment interactional behaviour in lessons with the "top" reading group and the "bottom" reading group that occurred on one day. McDermott found that the low-ability group children spent less time engaged in reading than did the high-ability group children, partly because their turn-taking procedures took time away from the instructional task. He also noticed that low-ability groups experienced more interruptions by other class members because the teacher tolerated more interruptions from other students in the class (i.e. outside the group) during lessons. He concluded that the agendas of high-ability and low-ability groups differed, with the low-ability students and the teacher wanting to avoid the frustration and embarrassment associated with getting through the lesson.

Eder and colleagues reported similar phenomena (Eder, 1981, 1982a, 1982b, 1983, 1995; Eder & Felmlee, 1984). Eder observed a first-grade (Year 2) classroom in the United States three days a week for an entire school year, and videotaped 32 reading group lessons in the fall and spring. She conducted sociolinguistic analyses of the videotaped interactions. As in other studies reported earlier, she found that the instruction of low-ability group members was characterised by a greater number of interruptions, more off-task behaviour, and a greater amount of time spent by the teacher managing student behaviour and attention. However, what was interesting was that she found that the teacher actually encouraged interruptions of oral reading turns by other students (what are termed 'call-outs') in low-ability groups but not in high-ability groups. Eder suggested that, since reading turns in low-ability groups tended to be longer and filled with more pauses, the teacher implicitly encouraged call-outs as a strategy for maintaining children's interest. Students were being socialised into "different communicative norms depending on their assigned group level" (Eder, 1982a, p. 261). Using quantitative modelling of attention shifts, combined with qualitative analysis of lesson transcripts, Eder and Felmlee (1984) also showed that students as well as the teacher aided the development of group attention norms and that these norms developed over time, becoming stronger in the spring than in the fall (see also Felmlee & Eder, 1983). Felmlee, Eder, and Tsui (1985) also documented peer diffusion of inattention in these groups. Distracting, inattentive behaviour on the part of one student increased the rate at which other students became inattentive. This 'contagion effect' was stronger in low- than in high-ability groups and it held up even after controlling for individual characteristics of group members.

Finally, Collins (1986) also obtained evidence that groups develop subcultures or norms of behaviour that differentially support learning. Collins analysed 16 videotapes of high- and low-ability reading groups in one first-grade (Year 2) classroom in the United States. Lessons were taught by the classroom teacher or the teacher's aide at the beginning, middle, and end of the year. As with previous studies, he found differential instructional strategies used with low- and high-ability groups (with a relative emphasis on decoding the text versus reading for meaning). However, Collins conducted a careful sociolinguistic analysis of the exchanges between students and teachers to look at what conceptions of the task of reading might be formed by the differing emphases, how they were formed, and how they were manifest in the reading lessons. The main focus of the analysis was on the mutual influences of children's reading-aloud style, defined in terms of prosodic features (that is, the features of the children's speech rhythms and intonations), and teachers' correction strategies. Children in low-ability groups segmented a text intonationally in word-by-word fashion when they read aloud, as if they were operating from a conception of reading as simply pronunciation. Teachers' corrections, in turn, focused on low-level linguistic information about letter-sound correspondences and words. By contrast, children in high-ability groups read texts with some of the intonational features of fluent, adult reading, as if operating from a conception of reading as making meaning. Teachers corrected their errors (sometimes the same sort of errors as those made by low-ability group children) with information about clauses, sentences, expressive intonation, and textual inference. What was interesting was that Collins found that the children's reading-aloud styles influenced the teachers' conception of their reading abilities, and the teachers' corrections, in turn, influenced the students' conceptions of the task (a similar phenomenon was noted in a study by Bondy, 1990). In the early lessons, teachers' expectations helped to produce students' conceptions of the reading task; whereas in later lessons, students' conceptions reinforced the teachers' expectations. Examination of the history of the teachers' correction strategies indicated that they were being socialised to respond differently in the two groups.

In summary, naturalistic research has documented a litany of differential learning experiences that might disadvantage students in low- as opposed to high-ability groups. The more fine-grained analyses just reviewed show that students engage with lessons in these groups according to 'norms of behaviour' that are built up, over time, through reciprocal teacher-student influences. What these findings suggest is that the nature of the differential learning experiences that children encounter in ability groups is more complex than a simple shift in teaching strategy may easily remedy (Cook-Gumperz, 1986). Hence, even if grouping practices and instruction are theoretically and empirically distinct factors, as Dreeben and others have pointed out, the cycle of reciprocal influences between teacher and students may be so strong as to militate against a shift in teaching strategy (cf. Cazden, 1988). Of course, because the naturalistic studies do not include outcome measures, it is impossible to judge the consequences of these differential experiences for students' learning. It would be expected that children in lower-ability groups would perform less well than children in higher-ability groups, since they are less proficient in their entry-level abilities. What is important is whether these experiences contribute differential outcomes over and above those due to initial differences.

3.2 Grouping and Mixing Students by Ethnicity and Gender

A small amount of research has investigated the effect of group composition on the basis of student ethnicity and gender (socio-economic status has been investigated only so far as it may be correlated with ethnicity). Most of this research has focused on peer-directed, cooperative small groups where the intention is to form groups heterogeneously to reflect the diversity of students' backgrounds and abilities. With some exceptions, these studies show that students' interaction and learning are "shaped by a combination of their own characteristics and those of the group they are in" (Webb & Palincsar, 1996, p. 858). Depending on the composition of the group, students' ethnic background and gender may serve as 'diffuse status characteristics' that influence interaction and

48

learning in cooperative small groups, especially where group members do not know one another or have no other basis for judging one another's competence on the task. These status characteristics determine students' relative influence in the group.

The warrant for these claims rests with *expectation states theory* (Berger, Cohen, & Zelditch, 1972). This is a general sociological theory that attempts to explain how the status characteristics of group members become the basis for expected competence in group tasks. Status characteristics are socially evaluated attributes of individuals for which it is generally agreed that one state (of the attribute) has higher status than another state. Some status characteristics are distinctions that are based on perceived ability to perform a specific task such as the ability to read – these are called *specific status characteristics*. Others are general social distinctions such as ethnicity or gender – these are called *diffuse status characteristics*. Expectation states theory claims that when a group is faced with a collective task, participants look for ways to judge the usefulness of their own contributions and those of others in the group. In the absence of direct information, they use status characteristics to make this judgement even if these charactertistics have no direct relevance to the task. Hence, those participants with higher status are expected to have greater competence at the task than those with lower status. As a result, the higher-status participants will be more active and influential than the lower-status individuals in the group task. Research, mostly in laboratory settings, provides extensive support for this theory (Berger, Rosenholtz, & Zelditch, 1980).

There is clear evidence that ethnicity may serve as a status characteristic in groups of mixed-ethnic composition. In North America, white students tend to be more active and verbal, whereas minority students tend to be more reticent and to participate less in mixed-ethnicity groups (Cohen, 1982a). In a series of studies of children playing a simple board game that required collective decisions, whites were found to be dominant over African-Americans (Cohen, 1972), Latinos (Rosenholtz & Cohen, 1984), and Canadian Indians (Cook, 1974). In Israel, Jews of Western origin were found to be dominant over Jews of Middle-Eastern origin (Cohen & Sharan, 1980). Also in Israel, in a study of classroom discussion in mixed-ethnic groups, Western Jews contributed more to the discussions than did Middle-Eastern Jews. These demonstrated inequalities in participation are probably linked to learning outcomes. Cohen (1984) showed that the status of a student was correlated with interaction in the group which, in turn, was correlated with learning, and that these relationships held up even after controlling for students' entry-level abilities.

There is evidence also that the relative dominance of high- and low-status students in small groups can be altered. Cohen and colleagues (Cohen, 1973; Cohen, Lockheed, & Lohman, 1976; Cohen & Roper, 1972) describe a series of 'expectation training' experiments in which they influenced status judgements of junior high school students. They did this by teaching low-status, African-American students how to perform academic and non-academic tasks and by having the low-status students teach the high-status, white students how to perform the task. The latter manipulation was important because it altered the white students' perceptions of the competence of the African-American students. As a result, the high- and low-status students showed equal rates of participation in mixed-ethnicity groups. These results were obtained in laboratory settings as well as in a classroom setting over a three-week period with a cooperative curriculum. Webb and Farivar (Webb & Farivar, 1994) report similar results from teaching Latino and African-American students' academic helping skills for use in cooperative small groups in Year 8 maths classes. Although not premised solely on expectation states theory, Webb and Farivar documented increased rates of participation and associated effects on learning at least for some of the minority students (the effects of the training were greater for one teacher than for another).

The evidence is less clear on whether gender serves as a status characteristic in mixed-gender groups. Lockheed, Harris, and Nemceff (1983) conducted a laboratory study of primary school children assigned to four-member teams of two boys and two girls. The boys were more likely to be perceived as leaders and to be perceived as having better ideas than girls, but these perceptions did not transfer to the activity or relative influence of the children working together on a board game. By contrast, Lockheed and Hall (1976) reported that 15- and 16-year-old boys were more active and influential than girls of the same age in groups engaged in the same task. Lockheed et al. speculated that gender may not yet be defined as a diffuse status characteristic until adolescence, "the seeds of which may be evident in the perceptions but are not evident in the behavior of fourth and fifth graders" (p. 888).

Webb (1984b) found that effects of gender may depend on the precise composition of mixed-gender groups. She compared the interaction and achievement of 77 Year 8 and 9 students in three kinds of mixed-gender groups: two girls and two boys, several girls and one boy, and several boys and one girl. Students worked in these cooperative small groups on maths activities for two weeks. She found that girls and boys in the balanced-sex group showed similar patterns of interaction and similar amounts of learning. By contrast, girls suffered in both the majority-girls and the majority-boys groups. In the majority-girls groups, the girls directed most of their requests for help to the boy, but he tended not to respond appropriately to their requests. In the majority-boys groups, the boys simply ignored the girl. In both cases, the breakdown in interaction impeded the girls' learning. These results suggest that gender becomes salient as a status characteristic only in majority mixed-gender groups; by making gender less salient in balanced groups, it may be possible to reduce inequalities in group interaction and learning (of course, single-gender groups may also be a solution to problems of imbalance in interaction between boys and girls).

Complicating the issue further, the function of gender as a status characteristic in peer interactions seems to vary across ethnic groups. Webb and Kenderski (1985) conducted a study very similar to Webb's (1984b) study but with mostly low-achieving African-American and Latino students (as compared to mostly white students). They found no significant differences between girls and boys on any interaction or outcome measures, regardless of the gender composition of the group. Corroborating evidence for this notion comes from an observational study by Grant (1986) of 15 primary school classrooms in the United States (cited in Webb & Palincsar, 1996). Grant noted that among white students, boys tended to dominate interactions with girls, whereas interactions among African-American students at all year levels tended to be more egalitarian. These findings suggest that ethnic background moderates the effects of gender on interaction and learning in mixed-gender groups.

In summary, there is good theory and evidence, both correlational and experimental, of compositional effects associated with the ethnic make-up of mixed-ethnic groups. Students' ethnic backgrounds determine their relative status within the group and therefore their interaction and learning. We do not know much about the strength of this effect, although there are clues in Cohen's (1984) and Webb and Farivar's (1994) work that the nature and extent of students' participation in a group is also greatly influenced by the teacher. Cohen, for example, notes that students' perceived status and entry-level abilities account for only eight percent of the variance in students' interaction (and, by default, that most of the remaining variance must be due to the teacher). The evidence is less clear on whether there are compositional effects associated with the gender make-up of groups. Certainly, there appear to be differences in the perceived status of boys and girls in mixed-gender groups, but these differences do not always translate into differences in interaction and learning. There may be developmental trends in the function of gender as a status characteristic, and the salience of gender may vary according to the mix of boys and girls in the group. As well, the effects of gender seem to interact with ethnicity.

3.3 Group Size

Very few studies have examined systematically the relationship between group size and learning outcomes. By and large, there seems to be some form of negative relationship, either linear or non-linear, between the number of students in a group and learning outcomes. In a review of research on small groups generally, Levine and Moreland (1990) comment that "as a group grows larger, it also changes in other ways, generally for the worse. People who belong to larger groups are less satisfied..., participate less often, and are less likely to cooperate with one another" (p. 593).

In Lou et al.'s (1996) analysis of grouped versus ungrouped classes, there were 92 contrasts that provided information on the effect of group size. A summary of the results showed that group size was significantly related to learning outcomes, although the relationship was non-linear. Pairs learned significantly more than students in ungrouped classes (mean effect size = .15); the optimal group size for learning, when compared to ungrouped classes, seemed to be three to four students (mean effect size = .22); and groups of five to seven students and groups of eight to 10 students did not learn significantly more than students from ungrouped classes (mean effect sizes –.02 and .11, respectively). It must be remembered that the 'grouped' classes in this analysis included both homogeneous and heterogeneous groups.

The best statement that can probably be made is that the effect of group size depends on the task set for the group. Bossert, Barnett, and Filby (1984), Kutnick (1994), and Steiner (1972) have all proposed typologies of group tasks and tried to develop conceptual frameworks that specify the interactions between group size and task as far as their effect on learning outcomes is concerned. These conceptual frameworks are still underdeveloped and we are far from having a good understanding of the 'best fit' between size and academic task characteristics. Nevertheless, some trends can be discerned from this literature.

In tasks where students' roles are not differentiated and they are not dependent on each other to accomplish the task – what Bossert et al. (1984) call *mechanistic grouping structures* (e.g., reading groups) – it seems that students are better off in small groups. In these situations, student attention and achievement are more dependent on teacher actions and, as noted by Bossert et al., "students in small groups receive more individualized assistance, more positive feedback and are exposed to a wider variety of materials than students who receive instruction in large groups" (p. 45). For example, a study by Peterson (1981) showed that high- and low-achieving students in small groups participated more frequently in lessons and retained more of the information than did similar students in large groups. Similarly, Sorensen and Hallinan's (1986) study of reading groups, reported earlier, found a strong negative relationship between the size of ability groups and reading achievement. This effect, though, was confounded with the homogeneity of the groups. Small groups tend to be more homogeneous than large ones (Steiner, 1972), and in Sorensen and Hallinan's study the small homogeneous groups contributed more to learning than did the larger, more heterogeneous groups.

In tasks where students work interdependently, and they may have specialised roles, to accomplish the task – what Bossert et al. (1984) call *organic structures* (e.g., cooperative or collaborative groups tasks) – interaction and learning involving all group members is more likely in small groups than in large groups. In pairs, it would be difficult for students to ignore one another's questions; in larger groups, there is more chance of 'social loafing', where students can shirk responsibility for helping others (Webb, 1989b). Most cooperative learning methods recommend using four-person groups, but the evidence on students' interaction and learning in groups of different sizes is equivocal. On the one hand, there is evidence that students in pairs and groups of three show similar patterns of interaction (Guntermann & Tovar, 1987). On the other hand, research has shown that students in three-person groups often ignore group members' questions (Webb, 1984b), but students in pairs seldom do (Webb, Ender, & Lewis, 1986). However, students in both configurations interact more than students in groups of four (Trowbridge & Durnin, 1984a; Trowbridge & Durnin, 1984b). None of the differences in patterns of interaction have translated into differences in learning outcomes.

Were it not for the finding of Lou et al., (1996), research on the effects of group size on learning could be summarised by saying 'smaller is better'. Taking into account all available evidence, the best generalisation that can be made is that smaller is better, but it depends somewhat on the nature of the task students are set. Smaller groups mean greater student involvement, but in some tasks, the nature of students' involvement may be restricted by the reduction in heterogeneity that the smaller size generally entails.

3.4 Informal Grouping of Students

This section focuses on peer effects operating in informal groups and their influence on learning outcomes. Informal groups are defined as *naturally* formed groups where students choose their own activities and partners, as opposed to formal learning groups whose composition and activities are *determined* by teachers or adults. Peer interactions within informal groups can be for academic or social purposes. One would logically expect that peer interactions for academic purposes would have more direct impact on achievement than peer interactions for social purposes. In this section, the contexts in which peer interactions in informal groups occur are limited to those related to informal talk in the classroom, extracurricular activities, and activities that occur during playtime and lunchtime.

3.4.1 Informal talk

Peer interactions take place not only in the official world of the teacher's agenda but also in the unofficial world of the peer culture (Alton-Lee, Nuthall, & Patrick, 1993). Analyses of student classroom utterances by Alton-Lee et al. (1993) revealed that 80% of the total utterances recorded during a 36-minute teacher-directed lesson was informal talk between students. A significant proportion of the informal talk involved the students' personal and hidden verbal responses to what was going on in the classroom. Students were often found discussing spontaneously with their neighbours what the teacher had just said. This often occurred when students became frustrated with the teacher because he or she did not respond to their hand signal; in such cases, the students would simply tell their neighbour the answer the teacher was failing to hear. These spontaneous interactions between peers can be thought of as transient "informal groupings" of students that exist for the life of a lesson.

Most of this informal talk between students (often dismissed as 'off-task behaviour') turns out to be conducive to learning. Alton-Lee and Nuthall pre- and post-tested students on content covered in curriculum units in science and social studies, and they collected extensive video, audio and other recordings of all students' learning experiences that might relate to ideas and concepts they had learned and not learned in the units (see Alton-Lee, Nuthall, & Patrick, 1993). They found that informal talk between students played a large role in fostering students' learning and memory for the ideas and concepts, no matter whether it was during teacher-directed lessons or independent activity time. According to Alton-Lee and Nuthall (1993), students' content-related informal talk afforded students opportunities for mutual support in their acquisition of new knowledge. Peers helped each other in creating associative links to their existing knowledge, evaluating the truth of their emerging understandings, and elaborating the content--yet all the time avoiding being seen by the teacher as contravening the rules of order during teacher-directed lessons (Nuthall & Alton-Lee, 1993). Peers also helped each other by exchanging tangible resources (e.g., felt-tipped pens, rulers, erasers) or intangible resources (e.g., critical and procedural information) in order to complete a learning task (Alton-Lee, 1984). The ability to gain access to these peer resources was also correlated with learning. Movement and spontaneous talk in classroom were related to learning because they optimised access to peer resources (Alton-Lee & Nuthall, 1990; Nuthall & Alton-Lee, 1990, 1992).

On the other hand, some of the classroom talk between students was found deleterious to learning. Alton-Lee, Nuthall, and Patrick (1987) found evidence of serious put-downs, teasing, blackmail and racial abuse during the hidden, private student-student interaction. Alton-Lee et al. reported a negative correlation between rates of giving or receiving derogatory remarks and learning of the intended curriculum.. When a student was giving or receiving derogatory remarks, he or she was less likely to learn and remember the curriculum content (Alton-Lee & Nuthall, 1990).

3.4.2 Extracurricular activities

Informal groupings of students also take place outside the classroom when students participate in extracurricular activities. These include athletic and non-athletic activities. Many non-athletic extracurricular activities are actually closely related to the school curriculum (e.g., science, maths, geography, astronomy, drama or foreign language clubs). It is reasonable to expect that peer interactions in these group activities should enhance classroom learning. Additionally, research findings suggest that peer interactions in other activities such as athletics, though having little connection to the curriculum, may also be beneficial. For both sets of activities, there is research that suggests participation is associated with better academic achievement, improved educational aspirations and attainment, and a lower dropout rate.

Before considering this research, it is worth noting that school size seems to moderate students' participation in extracurricular activities. Many studies have demonstrated an inverse relationship between participation and school size (Cotton, 1996; Fowler, 1995; Grabe, 1976, 1981; Gump & Friesen, 1964; McNeal, 1999; Stockard & Mayberry, 1992; Wicker, 1968). In small schools, individuals are needed to participate in order to populate teams, offices, and clubs. So even shy and less able students are encouraged to participate and made to feel they belong. By contrast, in large schools, a greater proportion of students do not participate in extracurricular activities because they are not needed to fill the available slots (Cotton, 1996).

Substantial research findings indicate that students' participation in extracurricular activities, including both athletic and non-athletic activities, is positively correlated with academic performance (Camp, 1990; Eidsomore, 1964; Gerber, 1996; Schafer & Armer, 1968; Steinberg, Brown, Cider, Kaczmarek, & Lazzaro, 1988; Sweet, 1986). For example, Sweet (1986) reported findings from a United States study that showed participants in extracurricular activities achieved higher grade-point averages than did non-participants. Gerber (1996) found the amount of participation of Year 9 African-American and white students was positively related to both groups' academic achievement. Cooper et al. (1999), in a study using a sample of 424 Year 7 through Year 13 students, examined the relationship between achievement and five after-school activities, namely, homework, TV viewing, extracurricular activities, other types of structured after-school groups, and jobs. They reported that more time in extracurricular activities and other structured groups and less time in jobs and TV viewing were associated with higher test scores and class grades. Using data set from the High School and Beyond (HSB) sophomore cohort, Camp (1990) tested a causal model of the relationship between level of student participation in extracurricular activities and academic achievement. The model accounted for the effects of individual students' gender, family backgrounds, academic abilities, and other competing time-use habits. Findings suggested that student participation enhanced academic achievement even after accounting for individual factors.

However, some studies report a curvilinear relationship between participation and achievement. For example, Cooper et al (1999) found an inverted-U relationship suggesting that extracurricular activities may become detrimental to achievement. Negative effects resulted if students' identification with an activity became so strong as to displace the broader school identity, or if the time investment was so great that it left little time for other out-of-school, academically related activities (e.g., homework).

Extracurricular participation has also been found to be positively correlated with higher educational aspirations and attainment, though only for boys. A number of studies have found that boys from lower socio-economic families who participated in athletics tended to have high educational aspirations (Otto, 1976; Rehberg & Schafer, 1968; Spady, 1970; Spreitzer & Pugh, 1973). Spady (1970, 1971) argued that participation in athletics was likely to enhance students' self-perceived status among peers and that this higher self-perceived peer status increased their educational aspirations (as measured by years of formal schooling). Otto and Alwin (1977), on the other hand, reported that a proportion of the total effect of athletic participation on educational attainment was mediated by significant others' influence substantially more than by self-perceived status. Other evidence from studies using causal modelling techniques, again for boys, indicates that the relationships between student participation and educational attainment are independent of obvious moderator variables such as individual socio-economic status and academic ability (Hanks & Eckland, 1976; Otto, 1975, 1976; Otto & Alwin, 1977) after controlling for moderator variables.

Two longitudinal studies found participation in extracurricular activities can also improve school dropout rate (Beauregard & Ouellet, 1995; Mahoney & Cairns, 1997). Mahoney & Cairn (1997) examined the relationship between participation in extracurricular activities and school dropout rate in a longitudinal study of 392 adolescents in the United States from Year 8 to Year 13.

Their results showed that the dropout rate among students identified as at risk for leaving school was lower for students who took part in the extracurricular activities. Mahoney and Cairn suggested that, for students whose prior commitment to the school and its values was marginal, participation in extracurricular activities provided an opportunity to create positive and voluntary connections to the educational institution. Posner and Vandell (1999), in a longitudinal study of low-income urban children from Year 4 to Year 6 found that participation in after-school activities were particularly beneficial for at-risk adolescents. Marsh (1992) suggested that participation in extracurricular activities might increase a student's investment in school, which may promote better academic attitudes and habits. Beauregard and Ouellet (1995) developed and evaluated a prevention program centred around extracurricular activities for five Year 9 students who were potential dropouts. Results showed that the program was successful. After the program, all participants showed improvement on measures of self-esteem, perception of the school environment, motivation and class attendance, and three of the participants showed improvement in their school marks.

Holland and Andre (1987) note that it may not be participation *per se* that influences desirable outcomes, but what happens as a result of the participation. Participation "may lead adolescents to acquire new skills (e.g., organizational, planning, time management), to develop or strengthen particular attitudes (e.g., discipline, motivation), or to receive social awards that influence personality characteristics" (p.447). Similarly, McNeal (1999) posits that student involvement in extracurricular activities is associated with increased levels of human capital (e.g., skills, years of schooling completed, and levels of achievement), cultural capital (e.g., specific attitudes and values), and social capital (e.g., extended sets of social relationships and networks).

Holland and Andre (1987) also caution that findings from the correlational and cross-sectional studies of participation in extracurricular activities need to be interpreted carefully. They note that participants and non-participants usually select themselves into or out of extracurricular activities; hence, pre-existing ability, personality and social differences between participants and non-participants may account for some of the observed correlations unless effects of these variables are adequately controlled (Holland & Andre, 1987). There is not much literature that can substantially delineate the causal sequence between student background, context, process, and outcome variables.

3.4.3 Playtime and lunchtime activities

Informal groupings that take place during playtime or lunchtime are mainly for social purposes - to play, to socialise, and to relax. Being part of an interactive peer group is important from a developmental perspective, and all children need to experience and have the opportunity to participate in variety of social relationships to develop social competence (Richardson & Schwartz, 1998). The social settings during playtime and lunchtime provide students with such opportunities. However, friends and peers on the playground can have positive or negative influences on students' attitudes towards learning and their academic achievement. These influences occur through the mediating effects of group norms, peer acceptance, friendship quality and stability, and individual students' level of social competence.

Group norms develop as students influence and are influenced by their friends' characteristics. Over time, these reciprocal influences lead to an increase in the similarity between friends (Steinberg et al., 1988). Many researchers have looked for evidence of friends' influence on each another by looking at correlations among friends on various attributes (Berndt, Hawkins, & Jiao, 1999). For example, Ide, Parkerson, Haertel and Walberg (1981) conducted a meta-analysis of 10 studies that examined the degree of similarity in academic aspirations and achievements between self and peers. They reported that the average correlation between individuals and comparison peers on academic aspirations and achievements measures was .24. Correlations were stronger among older students, in mixed gender samples, and in urban school samples. Of course, a drawback of the correlational studies is that similarities between self and peer affiliates can also be explained by pre-existing conditions or concurrent external factors, without the actual involvement of direct influences from peers. Individual students may select each other as friends because they are already similar in interest or proclivity. Influences from outside the relationship may also result in friends becoming more like one another (Urberg, 1999).

Different peer groups may have different group norms and these may have different effects. Some research findings indicate that peer group norms encourage individual students to work hard to get good grades (Brown, Clasen, & Eicher, 1986), whereas other studies report that peers contribute to a lack of effort and interest in schoolwork (Bishop, 1989; Goodlad, 1984). In a recent national longitudinal study, Chen (1997) investigated positive and negative peer influences on adolescents and found that adolescents who valued friends who cared about school did better in school, and adolescents who valued friends who were considered delinquent were more likely to be truant. Therefore, empirical evidence shows that peer effects on achievement depend on the attitudes and values of the peer with whom students spend most of their time (Epstein, 1983). If peers have little motivation to achieve in school, students' own motivation to achieve is likely to decrease over time; if peers have a high level of achievement motivation, students' motivation is likely to increase. Hallinan and Sorensen (1985) argue that streaming may exacerbate negative peer influences and racial tensions because students tend to choose their friends from among those in their assigned stream. Ianni (1989) notes that this friendship selection pattern may reinforce poor study habits and antisocial behavior among lower stream students.

Peer acceptance and the feeling of belonging to a group have also been found related to academic motivation and performance (Mpofu, 1997; Wentzel & Asher, 1995; Wentzel & Caldwell, 1997). Being rejected or feeling alienated by peers is related to lower levels of interest in school (Wehlage, Rutter, Smith, Lesko, & Fernandez, 1989; Wentzel & Asher, 1995) and school dropout (Hymel, Comfort, Schonert-Reichl, & McDougall, 1996). Ethnic minority students are more likely to encounter conflicts between peer group norms and achievement (Fordham & Ogbu, 1986; Pena, 1997). For example, many African American students view academic success as a form of 'acting white,' so peer pressure reduces their levels of effort and performance. Students in their early adolescence are particularly susceptible to peer pressure for fear of rejection or ridicule (Berndt, 1979). For adolescents whose friendship groups hold anti-academic or antisocial norms, peer pressure may pose dilemmas (Phelan, Davidson, & Cao, 1991). Emotional distress has been linked consistently to peer rejection and lack of peer support during this stage of development (Harter, 1990; Hogue & Steinberg, 1995).

Although students may need to struggle against the norms of their friends or their group, Steinberg et al. (1988) note that students have a choice of peers with whom to associate, and that most students are able to align themselves with peers who share their academic interests and aspirations. They also note that the freedom to select like-minded friends and to change associates as one's interests change can diminish the power of peer influences on students. Urberg (1999) suggests that all children and adolescents may not be equally likely to modify their behaviour to conform to their friends in order to gain peer acceptance. There appear to be demographic, individual, family, and relationship-specific variables that predict susceptibility. For example, studies of parents and peers find evidence that parents can influence their children to a much greater extent than peers can (Youniss & Smollar, 1989). A study of children without friends in middle school suggested that being liked by teachers might counteract whatever the negative effects of peer rejection might be on children's adjustment at school (Wentzel & Asher, 1995). Moreover, peer group membership tends to change frequently, suggesting that influences from a particular group might also be fairly transient (Wentzel, 1999).

The extent of peer influence may also depend on the quality and stability of friendships. Berndt et al. (1999) reported that Year 8 students' behavioural problems increased greatly when they had stable friendships with peers who had behavioural problems. It seems that very stable friendships magnified the influence of the misbehaving friends. The students' behaviour improved when they ended friendships with the misbehaving peers. However, two other studies (Berndt & Keefe, 1995; Urberg, Degirmencioglu, & Pilgrim, 1997) found no effects of stability of the relationship. Urberg (1999) suggests that "behaviours such as academic achievement may be much more sensitive to the stability of the relationship, because of the relatively long, delay between the behavior and its reinforcement" (p. 8). He also suggests that relationships in which there is more support, companionship, and intimacy will be more influential than relationships in which there are fewer positive qualities or more conflict.

Individual student's social competence may also determine the influence of peers. Some students lack the social competence to deal with interpersonal conflict in the playground and resort to physical or verbal violence. As a result, these students may find it difficult to develop and maintain friendship or to be accepted by their peers. Interpersonal conflict among children is very common in school. In a New Zealand study, Smith, Inder, and Ratcliff (1995) observed 102 primary school children for 20 minutes each in their classrooms and playgrounds to look at the incidence, context and nature of student conflict. They found that, on average, a child was involved in some type of conflict about six times an hour. If students lack social competence to resolve conflicts, playground activities may leave students with negative emotional feelings and low self-esteem (cf. Boulton & Smith, 1994). These, in turn, can lead to dropping out of school (Parker & Asher, 1987). Rigby and Slee (1993) found that low-self esteem and low levels of happiness were correlated with the frequency of being bullied as reported by a sample of 1162 secondary students, even after controlling for individual age and gender. In the 1996-1997 evaluation of truancy in New Zealand schools, McAlpine et al. (1998) reported that "the unhappiness created by being teased, bullied, and taunted had a marked impact on the decision to truant. The student often had other things going on in his or her life but the lack of support within the school setting, and the explicit harassment by other students facilitated the transition out of school. Friendship is a critical element in the retention rate of students returning to school" (p. 88).

In summary, peer interactions in informal groups can exert both positive and negative influences on students' social, emotional and academic development. The evidence on informal talk indicates that there are compositional effects although the usual experimental and statistical controls for individual factors, normally necessary to identify such effects, are absent in this research. The clever melding of a social-constructivist perspective with a cognitive model of information processing, and the detailed analysis of student talk in relation to learning outcomes, strongly point to effects that cannot be explained by individual characteristics of students alone. The evidence on extracurricular activities also suggests compositional effects. There are a growing number of more sophisticated studies showing that participation in extracurricular activities carries benefits for students' academic achievement, educational aspirations and attainment, and staying in school, even after controlling for pre-existing differences among students. The evidence on playtime and lunchtime activities is less clear. Based on available data, we are yet to be convinced that the influences of friends and peers go beyond what can explained by characteristics of individual students.

3.5 Implicating Peer Effects on Learning

As we have discussed so far, most research on small groups for instruction has examined either group characteristics in relation to outcomes, as summarised in the meta-analyses, or group characteristics in relation to social participation and instruction, as described in the naturalistic research. There is a need for research that examines group characteristics (e.g., heterogeneous versus homogeneous ability composition) in relation to students' social participation, academic task characteristics or instruction, and learning outcomes. Only if all four components are examined in the same study do we have a sound basis for evaluating peer effects on learning outcomes.

A small number of studies contain information on all four components and therefore provide a basis for evaluating peer effects in homogeneous and heterogeneous ability groups. In the area of homogeneous grouping, a handful of studies have found compositional effects of group membership (Anderson, Wilkinson, & Mason, 1991; Barr & Dreeben, 1983; Dreeben & Gamoran, 1986; Juel, 1990; Weinstein, 1976). All of these studies involved reading groups. These studies have shown that measures reflecting group membership, such as mean or nominal ability level of the group, add significantly to the prediction of individual students' performance, even after taking into account all possible individual characteristics of students (e.g., gender, ethnicity, socio-economic status, prior achievement). What this means is that the group measures contained something more than or different from the individual measures. These compositional effects may reflect measurement artifacts due to imperfect measurement of individual ability, differentially effective instructional practices on the part of the teachers, or differences in group cultures or norms of teacher and student behaviour (e.g., differences in norms for paying attention, differences in teachers' tolerance for call-outs).

Anderson, Wilkinson, and Mason (1991) found compositional effects even after steps were taken to obtain good measurement of individual students' abilities and the researchers held constant the nature of instruction. Anderson et al. conducted a microanalysis of small-group reading lessons in six Year 4 classes in the Midwestern United States. They used multiple indicators to obtain good estimates of individual students' abilities, and they scripted the lessons, including teachers' questions, and made sure all groups read the same texts. Although Anderson et al. reached no firm conclusion about the explanation for the effect of group membership, differential norms of behaviour was a strong candidate.

Norms of behaviour, established in reciprocal student-teacher interactions, are likely to be less conducive to learning in lower-ability groups as compared to higher-ability groups. Circumstantial evidence corroborating this conclusion comes from the work of Eder and Felmlee (1984) and Imai, Anderson, Wilkinson, and Yi (1992). Both these studies have shown a 'group effect' on inattention in groups of lower ability beyond that which can be accounted for by students' individual characteristics. Norms of behaviour that are not conducive to learning might help account for Lou et al.'s (1996) finding that low-ability students fare worse in homogeneous than in heterogeneous groups.

In the area of heterogeneous grouping, peer effects are easier to identify as these groups tend to be peer led rather than teacher led. The only major confounding factor to consider is academic task characteristics. A programme of research by Webb (for a review, see Webb, 1991) has investigated the verbal interaction and cognitive mechanisms mediating learning in cooperative small groups. These studies contain information on all four characteristics necessary to identify peer effects. Students in these groups were engaged in mathematical problem-solving and computer programming. A consistent finding from this research is that giving elaborated explanations facilitates the learning of the students giving the explanation, even after controlling for their entry-level characteristics (Webb, 1980b, 1980c, 1982a, 1984c, 1989b, 1995b; Webb & Kenderski, 1984; Webb & Kenderski, 1985). The most likely explanation for this is that giving elaborated explanations fosters cognitive restructuring and cognitive rehearsal on the part of the explainer. Another finding, though less consistent, is that receiving elaborated help that is timely and responsive to the needs of individual students benefits the students who ask for the help (Webb, 1980c, 1982a, 1992, 1993, 1995c). Receiving timely, relevant, and elaborated help may enable students to correct their misconceptions and may foster greater engagement and constructive problem-solving activity. More students benefit from giving and receiving help when the range of abilities within the group is not too wide (Webb, 1982b, 1984c; Webb & Cullian, 1983b; Webb & Kenderski, 1984).

These findings demonstrate peer effects in heterogeneous groups stem directly from verbal interactions between students of higher and lower ability. The nature of these interactions might also help account for the finding in Lou et al.'s (1996) meta-analysis that low-ability students derive more benefit in heterogeneous groups (in addition to the observation that norms of behaviour in homogenous low groups tend not to be conducive to learning). This may be because the low-ability students are receiving timely and elaborated help from their high-ability peers. The high-ability students may also benefit because they are the ones engaged in giving the elaborated help. By contrast, the medium-ability students perform poorly in heterogeneous groups because they participate little in the giving and receiving of help; they learn more in homogeneous groups.

Similarly, the effects of ethnicity and gender on learning in heterogeneous groups stem directly from verbal interactions among peers. As mentioned, depending on the composition of the group, a student's ethnicity or gender may serve as a status characteristic, particularly in situations where there is no sound basis for making judgements about another's ability. These characteristics then become 'proxies' for ability that determine a student's relative influence and learning in the group.

The negative relationship between group size and learning reflects fewer opportunities for learning in larger groups. This may arise from teachers having to divide their time among larger numbers of students, if the groups are teacher led, and/or greater inattention in groups of larger size (Sorenson & Hallinan, 1986). Imai, Anderson, Wilkinson, and Yi (1992) found a trend, albeit nonsignificant, for more inattention in groups of larger size. To the extent that norms for paying attention implicate peers, this negative effect of group size may represent a peer effect. Because these studies do not include information on all four variables needed to implicate peer effects, there is insufficient basis to make a judgement on this issue.

Finally, as with interactions in heterogeneous peer-led groups, interactions in informal groups readily implicate peers. The strong theory and careful data analysis underpinning Nuthall and Alton-Lee's (1993) work demonstrates clearly the cognitive benefits of informal talk. The most likely explanation for these benefits is that informal talk promotes cognitive restructuring, cognitive rehearsal, co-construction of ideas between peers, and perhaps internalisation of problem-solving strategies and activation of inert knowledge. These mechanisms are similar to those identified as mediating learning in Webb's (1991) heterogeneous peer-led groups. The benefits of participation in extracurricular activities also implicate peer group influences. The benefits of extracurricular activities probably come about by ambient mechanisms such social comparisons, enhancing social norms about the value of school, group socialisation, and promoting socio-emotional support among peers. The effects of playtime and lunchtime activities would also seem to implicate peers but the data are simply not good enough to make any claims on this point.

3.6 Relevance to New Zealand

Special mention must be made of the relevance of the overseas literature on grouping to schooling in New Zealand. There is some evidence from case studies that classroom cultures in New Zealand differ from those in the United States, at least in primary schools, and that different cultures or norms of behaviour may prevail in small-group settings (Wilkinson & Townsend, in press; Wylie, 1996). We deal first with New Zealand studies of homogeneous grouping, then with studies of heterogeneous grouping.

With respect to homogeneous groups, Wilkinson and Townsend (in press) conducted an intensive case study of grouping for reading in four New Zealand classrooms. They identified four teachers from four schools in Auckland who were regarded by fellow teachers and others as exemplifying 'best practice' in the teaching of reading. They interviewed the teachers every few weeks throughout the school year about their groups, reasons for their decisions regarding grouping, and their instructional practices associated with grouping. These interviews were supplemented with regular observations of classroom practices, analyses of documents supporting the teaching of reading, and information from other sources. Analyses of the interviews, observations, and teaching guides converged on the conclusion that a key factor contributing to the apparent success of grouping for reading in New Zealand primary schools – as judged by New Zealand's performance in international surveys of reading ability – was that teachers held a developmental notion of ability. That is, rather than assuming that ability was innate and unchangeable, teachers seemed to view ability as incremental and malleable. Hence, grouping arrangements operated within a classroom culture that stressed *continual development* of children's expertise rather than accommodation to fixed traits or abilities. The researchers also noted that groups were just one of a number of organisational arrangements teachers used. Wilkinson and Townsend concluded that ability grouping, at least in the classrooms they studied, could provide effective contexts for teaching lower-ability readers as well as higher-ability readers. Moreover, they suggested that, because of the somewhat greater flexibility of grouping in New Zealand

and the classroom culture in which groups operate, such contexts might provide minimal opportunities for differential norms of behaviour to take hold.

These conclusions compare with those of another New Zealand study. Wylie and Smith (1995) followed the progress of 30 children from 10 schools over the first three years of schooling in New Zealand. Using a descriptive case study approach, they reported on perceptions of progress through the eyes of the children, their teachers, and their parents. On the issue of within-class grouping, Wylie and Smith made this observation:

> *The conclusion from the American literature is that ability grouping is a significant shaper of experience, academic self-concept, and academic achievement (Rosenholtz & Simpson, 1984). This material from New Zealand children shows a different picture. The reasons for this are likely to be differences in classroom organization, assessment, and teacher orientation to progress* (p. 47).

They go on to explain these reasons – that children were in different groups for different curriculum areas, that groups were just one of several organisational arrangements, that there was a wide range of ability in most classes, and that assessment was largely curriculum based and non-comparative. They also point out that New Zealand teachers' perceptions of progress stem from a 'learning' as opposed to a 'performance' goal orientation (using the distinction made by Dweck, 1986, in the motivation literature). According to Wylie and Smith, New Zealand teachers, at least in the first three years of school, see children's ability as changeable through effort rather than as inherent and fixed, largely because of the curriculum and assessment context in which they work. These conclusions bear a striking similarity to those of Wilkinson and Townsend (in press).

These case studies suggest that homogeneous ability grouping may not have a negative shaping influence on students' learning in New Zealand classrooms. However, as with the naturalistic studies conducted overseas, these local studies document the relationships between group composition, children's participation, and the nature of instruction, but they provide no evidence of students' learning. Until such evidence is forthcoming, findings from these studies must be regarded as suggestive and in need of confirmation using different methods of inquiry.

With respect to heterogeneous grouping, Rzoska and Ward (1991) and Jacques, Wilton, and Townsend (1998) have conducted studies using cooperative small groups in New Zealand. However, we know of no local research on students' verbal interactions and learning in such groups along the lines of that conducted by Webb (1991) and her colleagues. Similarly, we know of no studies that have examined effects of the ethnic composition of cooperative small groups in New Zealand, nor do we know of any studies that have attempted to teach minority group students academic helping skills to promote interaction and learning in these groups (cf. Webb & Farivar, 1994). Hence, we do not know whether the conclusions on effects of ethnic composition of small groups, predicated on expectation states theory and supported by overseas studies, apply to Maori, Pacific Island, Pakeha, and other ethnic groups in New Zealand.

Grima (1999) conducted a study of the effects of gender composition on group performance in peer-led small groups. Her study was a doctoral thesis based on videotape data collected for the National Educational Monitoring Project (NEMP) (Crooks & Flockton, 1994). Grima analysed tapes of four-member groups working on tasks in three curriculum areas (science, language, and technology) at two year levels (Year 4 and Year 8). For each task, approximately 90 groups were selected randomly from the larger NEMP data set representing all combinations of boys and girls: all boys, all girls, three boys and one girl, three girls and one boy, or two boys and two girls. She carefully and exhaustively analysed students' interactions, group processes, and overall performance of the groups. Results showed no consistent differences between groups of varying gender composition, at any year level, for any task. However, there was a tendency for the minority student in the groups of three girls and one boy or three boys and one girl to contribute less than the other group members, and for these students to contribute less than their same-gender peers working in other groups. The lack of consistency in findings from this study reflects the inconsistency in findings on gender composition from overseas studies.

3.7 Conclusion

Research shows that there is a small advantage of forming students into groups for instruction as opposed to using whole-class instruction. This seems to be true particularly when class sizes are large (mean effect size = .35). Within groups, there is an advantage of homogenous over heterogeneous grouping though this result depends on the curriculum area and probably task. Forming students into homogeneous ability groups as opposed to using whole-class instruction carries with it benefits in the order of .23 of a standard deviation or .30 if we limit our analysis to the research syntheses using more stringent criteria for study inclusion. We note, however, that effect sizes can be most variable, that the studies are dated, and that the overall effects for some studies are small compared to most other school effects (see Table 1.2).

The evidence from studies of the effectiveness of ability grouping and from naturalistic studies of instruction and social participation in ability groups is contradictory. On the one hand, ability grouping seems to be effective for students of all ability levels. On the other hand, there is a risk that low-ability students' learning may suffer in homogeneous groups, not only from less instruction and less effective instruction but also from norms of behaviour that are not conducive to learning. The latter reflect a peer effect to the extent that peers contribute to a cycle of reciprocal teacher-student interactions that evolve over time.

How can this contradiction be resolved? One explanation is that the differential learning experiences of students in low-ability groups do not contribute to lower achievement beyond what would be expected from their entry-level abilities. Because the naturalistic studies do not include outcome measures, it is impossible to judge the consequences of these differential experiences for students' learning. It would be expected that children in lower-ability groups would perform less well than children in higher-ability groups since they are less proficient in their entry-level abilities. What is important is whether these experiences contribute differential outcomes over and above those due to initial differences. Another, explanation is that most of the effectiveness research has focused on maths groups whereas the naturalistic studies have focused on reading groups. Compared to maths, we know very little about the effectiveness of grouping for reading in terms of the learning outcomes for students of different ability.

Heterogeneous groups seem to be effective for students of higher- and lower-ability, presumably depending on curriculum area and academic task characteristics. In heterogeneous groups, students of higher- and lower-ability form a teacher-learner relationship (in narrow-range mixed-ability groups, medium-ability students can also fare well if they are perceived as being of higher ability). Depending on the composition of the group and other factors, students' ethnicity, and possibly gender, may also determine students' relative status and therefore their interaction and learning in the group. These effects are peer effects since they stem directly from verbal interactions among students of higher and lower ability or among students whose characteristics are perceived as proxies for ability. We do not know if the effects associated with ethnicity apply in New Zealand.

There seems to be some form of negative relationship between number of students in a group and learning outcomes though we do not know if this is a peer effect. This relationship may arise because teachers have to divide their time among larger numbers of students or because there is a lack of involvement of students in larger groups. Even if group size has a peer effect, students' interactions in large and small groups do not appear to be qualitatively different from those occurring in groups of varying composition (by ability, gender, or ethnicity). Hence, we continue to view group size as moderating variable that simply magnifies or attenuates the effects of group composition.

Peer interactions in informal groups inside and outside the classroom are associated with social and academic outcomes that may contribute to student learning. In the case of informal talk and participation in extracurricular activities, we are reasonably confident that there are peer effects.

However, in the case of playtime and lunchtime activities, we do not know if there are compositional effects, let alone peer effects, because the research has not adequately controlled for pre-existing personality and social differences between participants and non-participants, nor has it controlled for individual demographic characteristics.

3.8 Recommendations for Further Research

- There is need to resolve discrepancies in the findings from different meta-analyses of the effects of homogeneous, within-class ability grouping relative to whole-class instruction. In particular, there is need for research on the effectiveness of grouping for reading in terms of the learning outcomes for students overall, as well as for students of different ability.

- The findings on instruction and social participation in ability groups all come from studies conducted in North American classrooms; we need to know if membership in lower-ability groups in New Zealand classrooms also conveys negative consequences for students.

- There is need for research that examines group characteristics (e.g., heterogeneous versus homogeneous ability composition) in relation to students' social participation, academic task characteristics or instruction, and learning outcomes. Only if all four components are examined in the same study, do we have a sound basis for evaluating peer effects on learning outcomes.

- There is need for research on interaction and learning of students of different ethnic backgrounds in New Zealand (Maori, Pacific Island, Pakeha, Asian) as they work in cooperative tasks in groups of mixed-ethnic composition.

- There is need for more sophisticated, conceptually driven, research on students' participation in extracurricular activities and playtime and lunchtime activities and the contribution to students' social, emotional, and academic development. In particular, we need to understand the effects of informal groups during playtime and lunchtime activities. Research needs to delineate the causal connections between student background, context, process, and outcome variables.

CHAPTER 4

CLASS CONFIGURATIONS

This chapter is concerned with class configurations that could implicate peer effects. The major configurations discussed are streaming or 'tracking', class size, composite classes, and single-sex classes. Data with respect to the New Zealand context are presented, and the implications of overseas data for New Zealand are explored. The premise of studying peer influences at the class level is that instruction can be more appropriately matched according to the compositional grouping of students. For example, with respect to streaming, it is claimed – often as an extension of the aptitude-treatment interaction research (Cronbach & Snow, 1969) -- that more homogeneous groups allow teachers to change the nature and/or pace of instruction to better match the needs and abilities of the students. Although class size is not a compositional factor in the same way as between- or within-class grouping, it may have an effect in that there could be more peer interactions in classes of smaller than of larger size.

Because of variants in streaming practices and terminology, it is often difficult to derive estimates of the extent of streaming, particularly in New Zealand primary and intermediate schools. The majority of New Zealand secondary classes seem to be streamed. Werry (1987), in the Second International Mathematics Study, found that 66% of the 152 schools surveyed used streaming, 16% had some streaming, and 18% had no streaming. In the United States, it is often claimed that about 20% to 40% of intermediate schools assign students to all classes on the basis of ability, and a further 40% used some between-class grouping, primarily in reading and mathematics (Epstein & MacIver, 1990; Lounsbury & Clark, 1990; Wheelock, 1992). Data from the National Educational Longitudinal Study (NELS) of 25,000 students in nearly 1,000 schools show that an estimated 86% of public school students in United States middle and high schools are placed in streamed classes; for students in independent schools the estimate is slightly lower, at 71%. A problem with these estimates is that they are often based on self-report data, whereby students report whether they are in streamed classes. Students may not make the more refined analyses between configurations and could confuse within-class grouping with between-class grouping; thus, the estimates are likely to be upper limits. Approximately 80% of middle schools use streaming, although 36% of these are considering 'detracking' – that is, creating, or reverting to, unstreamed (untracked) classes (George & Shewey, 1994; Mills, 1998; Valentine, Clark, Irvin, Keefe, & Melton, 1993).

In those schools that do use between-class grouping, there is still much evidence of streaming within certain subjects. Loveless (1998) estimated that 39% of all United States schools have students streamed into three groups (high, middle, low) for all subjects, 18% have two groups for all subjects, 11% have three groups for some subjects and unstreamed for others, 10% have two groups for some subjects and unstreamed for others, seven percent have one subject streamed and unstreamed for others, and 14% have all unstreamed classes. Streaming is more likely to be used in high schools with rolls of more than 200 students, which is not surprising given that smaller schools are unlikely to have a sufficient number of classes to consider streaming. Secondary schools with more than 500 students are almost certainly to be streamed in the United States (Loveless, 1999b). Streaming is also common in areas with much 'bright flight', as it is seen as a way of holding on to parents and students who seek advantage from the perceived higher-achieving schools (Oakes, 1992).

The amount of data available in New Zealand on class size at the national level is small. The closest information relates to teacher-pupil ratios that, while providing some indication, can be quite different from class size (Finn & Achilles, 1999). In the late 1980s, the New Zealand Government introduced a 1:20 teacher-pupil ratio for junior classes (although teacher-pupil ratio is not the same as class size, the introduction of a reduced teacher-pupil ratio had the effect of reducing class size). After various evaluations and reviews (e.g., Renwick et al., 1989), the New Zealand State Services Commission (1991) ultimately advised that the 1:20 ratios be suspended, the effect of which was for ratios to increase. In 1994, a new edict was pronounced that teacher-pupil ratios should be 1:23 for

Years 1 to 3 of school. By 1996, 47% of teachers of new entrants and 24% of teachers in the junior school had class sizes of fewer than 20 pupils (Wylie, 1997). In 1998, the ratio of students to each teacher was 19.5 for primary, 14.9 for intermediate, and 15.9 for secondary schools (Ministry of Education, 1999a). In a study of 523 eight-year-olds attending 168 schools in New Zealand, Wylie, Thompson, and Lythe (1999) found that the median class size was 28, with the smallest at 13 and the largest at 40. There were smaller classes in private schools (64% of the children who attended private schools were in classes of less than 25, compared to 30% of those attending state schools). Classes were also smaller in schools with smaller proportions of Maori students (35% of the children whose schools had less than 15% Maori students were in classes of less than 25, compared to 25% with larger proportions of Maori students).

Three-quarters of primary school classes in New Zealand are composite classes (classes comprising students of more than one Year level). According to data collected from the 1990 IEA survey of reading literacy, 76% of classes comprising Year 5 students were composite classes (Wilkinson, 1998). The data from the 1994 Third International Maths and Science Study (TIMSS, another IEA survey) showed that almost 70% of classes comprising Year 4 or Year 5 students were composite classes (Chamberlain, 1997). These figures are high by international standards. In the IEA survey of reading literacy, only Portugal and Trinidad and Tobago had comparable percentages (72% and 73% respectively). Most countries reported having around 50% or fewer composite classes (Chamberlain, 1993).

4.1 Streaming

Class-based streaming or 'tracking' refers to a "sorting process whereby students are divided into categories so that they can be assigned in groups to various kinds of classes" (Oakes, 1985, p.3) (The term 'tracking' is used more in the United States and the term 'streaming,' in New Zealand, though the terms can be used interchangeably). There are many forms of streaming, although the fundamental concern relates to whether there is heterogeneous or homogeneous grouping. Such grouping is typically formed on the basis of ability or achievement, although students can be assigned on the basis of combinations of achievement, IQ, and teacher judgements to a stream within which all courses are taken. This is usually termed 'tracking' in the United States and 'streaming' in the United Kingdom and New Zealand. 'XYZ skill grouping' has been used to refer to students grouped together for purposes of instruction in, usually, three levels – high-, middle-, and low-streamed classes (Mosteller, Light, & Sachs, 1996). Another form is where students spend most of the day with a homogeneous group of students, but this group can differ according to subject; thus a student may be with the higher-streamed group for mathematics and in a middle group for English. This is sometimes referred to 'block scheduling', although this term has been used more recently for many other forms of scheduling. For example, a variant is that the students be blocked/tracked for some subjects (typically mathematics and English) and in heterogeneous classes for other subjects (e.g., social studies, physical education).

The 'Joplin Plan' is a more specific form of arranging homogeneous classes, usually in a specific subject. For example, imagine that there are students grouped according to age into Years 5, 6, and 7 and that the reading levels of these students ranges from Level 1 to Level 9. For the purposes of reading, however, students are grouped by reading level regardless of age. When the reading class is over, the students return to their original classes for other subjects. A further consideration is that streaming in the upper secondary schools often involves students undertaking different courses, whereas in the earlier Years it typically involves students taking the same subjects but the orientation of the instruction is intended to differ to match the differing ability levels. At the intermediate school level, it is more likely that students are streamed in some subjects (e.g., English and/or maths) and are in unstreamed classes for other subjects. There have been very few studies using upper secondary students, and thus the review will relate to students in Years 1 to 10.

There are, of course, multitudes of outcomes that can be used to address the question as to the effects of streaming. The outcomes can be broadly grouped into two: achievement effects, which include subject matter scores, or some overall composite achievement scores; and equity effects, which address the question as to whether the gains from streaming are uniformly distributed across various subgroups (e.g., minority versus majority students; high- versus medium-, versus low-streamed groups). Many of the studies relating to the equity outcomes also address concerns about whether there are differences in instructional pace and teaching methods moderated by subgroups and whether there is differential access into the streams on variables other than the avowed grouping variable (e.g., if social class influences access over and above achievement level).

The basic premise in favour of streaming is that it is a harmless meritocratic practice that creates homogeneous instruction groups based on students' prior achievement within a heterogeneous student population. Streaming decisions can represent the schools' efforts to use "educational structures and technologies to match students and courses in ways that both further societal social goals and accommodate individual differences" (Oakes & Guiton, 1995, p.5). Streaming permits students to make progress commensurate with their abilities and thus helps to maintain interest and incentives for students. It also aims to prepare students for productive work and to mirror the stratified world of work (and social class) they are about to enter. In streamed classes, slower students are less likely to be eclipsed by those much brighter than them, and brighter students are not bored or slowed down by the presence of less able students. Further reasons for the wide adoption of streaming are that it allows for more efficient scheduling and clearer opportunities for students to follow well-defined course sequences and that it enables schools to optimise teachers and teaching resources in a way that makes best use of teachers' subject specialities and allows for better educational decisions to be made regarding scarce teacher resources.

In contrast, the most cited negative effects are that streaming can often institutionalise students into a similar stream throughout their high school experience; remove the advantages of peers of varying ability assisting each other, particularly low-ability students being stimulated and encouraged by more able students; assign a stigma to the lower streams, which can operate to discourage these students; lead to teachers being less able, or not having time, to differentiate the work for different levels of ability; lead to teachers objecting to, or not wishing to, teach the lower-ability streams; lead to minority and lower-class students being more likely relegated to lower-ability streams; and mean that students in low streams receive a lower pace and lower quality of instruction. Slavin (1990) summarised the typical claims against streaming: "Because of the demoralisation, low expectations, and poor behavioral models, students in the low tracks are believed to be more prone to delinquency, absenteeism, dropout, and other social problems" (p. 473) (Crespo & Michelna, 1981; Gamoran, 1987a, 1987b; Wiatrowski, Hansell, Massey, & Wilson, 1982). Streaming is claimed to perpetuate social class and racial inequalities in the lower streams, and is often considered to be a major factor in the development (or maintenance) of elite and underclass groups in society (Persell, 1977; Rosenbaum, 1980).

Thus, the major benefits of streaming relate to more efficient adaptation of instruction to the needs of a diverse student population (with more reference to the advantages to higher-streamed students), whereas arguments opposed to streaming focus on the over-assignment of minorities and students from lower socio-economic groups to the lower streams and the detriment to low achievers, who receive a slower pace and lower quality of instruction (e.g., Gamoran, 1989; Oakes, 1985; Persell, 1977; Rosenbaum, 1980).

There are many major research approaches in the streaming literature. Meta-analysis involves statistically synthesising a large number of studies to estimate an overall effect size and then systematically investigating various critical moderators of this effect size. The major outcome in these studies typically relates to achievement. In regression-based studies, there is more control of the various factors that may moderate the overall conclusions. The most recent studies of streaming often use regression methods based on secondary analysis of large-scale national samples, such as NELS and the National Assessment of Educational Progress (NAEP). Comparison designs compare streamed and non-streamed classes, but, unless the groups are similar at the outset (often achieved by random assignment to the streamed and non-streamed classes), any pre-group differences can complicate the demonstration of later differences. Qualitative studies investigate, in detail, a smaller number of classes using ethnographic methods and often highlight equity and teaching-related outcomes.

4.1.1 Effectiveness studies: Experimental studies

The meta-analysis studies have summarised over 700 effects of streaming, covering a wide variety of schooling cultures and experiences, in most curriculum subjects, across all age ranges, and across most major educational outcomes. For example, Kulik and Kulik (1982a) synthesised 52 effect sizes based on 52 studies of streaming carried out in secondary schools. They selected empirical studies that met three criteria: they had to take place in secondary schools, they had to report on measured outcomes in both streamed and unstreamed classes, and they had to be free from crippling methodological flaws. They used only the average effect size from each study and thus lost many of the nuances and much of the detail that can be gained from meta-analyses. The overall effect size was .10. It is important to note that the Kulik and Kulik study included a number of studies of gifted and talented students, and it was these that contributed most to the positive effects, with an average effect size of .33. When these studies were removed, the average effect size was .02. Further, in the studies with random assignment to streamed or unstreamed groups, the effect was –.01, and when the same instructors were used, it was .07, compared to .14 when different teachers were used. The effect sizes reduced as the duration of the study increased (.20 for studies of 5–18 weeks duration; .11 for 19–36 weeks duration; and .00 for 37 or more weeks). The effects were greater (.20) at senior high school compared with junior high school (.05), and close to zero in maths (.05) and reading (.02). The size of the effect of streaming on self-concept was .01, but there were greater effects on attitudes towards subject matter – here, the effect size was .37, indicating that streaming has a more positive effect on student attitudes towards the subject being taught. The effect size for attitude towards schooling was lower (.09).

Kulik and Kulik (1984) reported a meta-analysis of the effects of streaming on primary students. They chose 31 studies that compared streamed versus unstreamed groups, and excluded studies of within-class grouping, rapid promotion, and non-graded schools (Table 4.1). On the basis of 23 studies, they reported an overall effect size for achievement of .19, and again found that the majority of the studies contributing to the positive effect size came from programmes designed specifically for gifted and talented students (.49), whereas, for programmes for more representative populations, the overall effect size was .07. The overall size of the effect on self-concept was .06. From our analyses of the data provided by Kulik and Kulik, the effect size for the lower primary Years 1 to 3 was .02 and for Years 4 to 6 was .26. The longer the implementation of the streaming, the smaller the effect size ($r = -.17$). (Note that, in their 1987 meta-analysis, Kulik and Kulik added the primary and secondary level studies together and added no new studies, which led to an overall average effect of .06.).

Slavin (1984a) criticised the Kulik and Kulik meta-analyses because they included too many studies with inadequate experimental controls (such as poor matching, non-random assignments of teachers). Marsh (1984a) also claimed that the analyses of self-esteem findings gave too little attention to interaction effects, whereby self-concept effects may occur within certain sub-dimensions of self and across different age groups. Hence, in a third study, Kulik and Kulik (1985) reported on a refined set of studies, with particular attention to self-concept outcomes. They identified 85 studies, 40 at the primary level and 45 at the secondary level. For each study, they coded 13 experimental design features to ascertain the possible moderator effects of the attributes of the study on the overall effect size. The overall effect size for achievement was .15, and again they reiterated the critical importance of the nature of the grouping. The effect size for gifted programmes (25 studies) was .33, for XYZ programmes it was .12, and for remedial programmes for slower learners it was .14. The effects for the XYZ programmes also differed by level of ability with high-ability groups having higher effects (.12), compared to middle-ability (.04) and lower-ability groups (.00). The effect on self-concept was .00. Kulik and Kulik concluded that "homogeneous grouping seemed to have little effect on the achievement and self-esteem of students at the middle level of ability…. It also appeared to us that XYZ program effects were more moderate than were effects in programs for special target groups" (pp. 6–7).

Table 4.1

Summary of 23 Studies Reviewed by Kulik and Kulik (1984)

Study	Year levels	Duration (months)	Effect size
Koontz (1961)	4	9	-.31
Barker-Lunn Study II (1970)	3	9	-.27
Daniels (1961)	1	36	-.24
Goldberg, Passow & Justman (1966)	5	18	-.13
Bremer (1958)	1	9	-.12
Johnston (1973)	1	9	-.03
Loomer (1962)	5	9	-.02
Barker & Lunn Study I (1970)	2	27	-.01
Hartill (1936)	5	5	.01
Flair (1964)	1	9	.04
Moses (1966)	5	4.5	.05
Breidenstine (1937)	3	27	.08
Morgenstern (1963)	4	27	.15
Cluff (1964)	4	18	.23
Provus (1960)	5	9	.27
Berkun, Swanson & Sawyer (1966)	4	9	.32
Borg (1964)	5	9	.32
Barthelmess (1932)	4	9	.38
Jones & McCall (1926)	5	18	.60
Atkinson & Connor (1963)	6	9	.61
Bell (1959)	5	9	.68
Luttrell (1959)	6	7	.70
McCall (1928)	3	18	.71

Slavin's (1987) synthesis used his 'best-evidence' method for undertaking a meta-analysis. Slavin's criteria included only comprehensive studies of streaming in primary schools, although he did not specify what 'comprehensive' meant. The methodological requirements included: streamed classes were compared to heterogeneously grouped control classes (and thus excluded studies that reported gains within one streamed class over time); achievement data from standardised achievement tests must be available; initial comparability of samples was established by random assignment, matching of classes, or matching of students within equivalent classes; streaming had to be in place for at least one semester; and at least three experimental and three control teachers had to be involved in the study. Like the Kuliks, Slavin calculated only one effect size per unique sample within a study. He located 14 studies of streamed primary classes, yielding 17 effect sizes, with an average of .00 (Table 4.2). Ten of these studies were common to the Kulik and Kulik (1984) meta-analysis, so it is not surprising that the results are similar. The overall effect size for reading was –.01, and for mathematics –.04. The effects were .03 for the high-streamed groups, .02 for the middle streams, and .02 for those in the low streams.

Table 4.2

Summary of 14 Primary School Studies Reviewed by Slavin (1987)

Study	Year levels	Duration (months)	Overall	Reading	Maths	Effect size High	Effect size Medium	Effect size Low
Cartwright & McIntosh (1972)	1-3	16	-.34	-.17	-.28			
Barker-Lunn (1970)	2-5	32	.00	.00	.00	.00	.00	.00
Borg (1965)	4-7	32	.00	.00	.00			
Goldberg, Passow, & Justman (1966)	5-6	16	na					
Hartill (1965)	5-6	2.5	.00	.05	.01	-.12	.00	.18
Bathelemess & Boyer (1932)	4-5	8	.21			.18	.22	.15
Tobin (1966)	2-6	16	.05	.13		.13	.03	-.46
Breidenstine (1936)	2-6		-.08					
Rankin, Andersen & Bergman (1936)	3-6	16	.05	.03	.07	.1	.03	.12
Daniels (1961)	2-5	28	-.26	-.25	-.27			
Bremer (1958)	1	8	-.1	-.1		-.24	.00	-.06
Loomer (1962)	4-6	8	-.04			-.02	.04	-.06
Flair (1964)	1	8	-.06	.03	-.14	.54	-.21	-.11
Morgenstern (1963)	4-6	24	.15	.17	.06	-.22	.15	.64

The results for streaming, Slavin claimed, are "surprisingly clear cut". He stated that "there is no support for the practice of assigning students to self-contained classes according to general ability or performance level... [and it is recommended that we] "avoid ability-grouped class assignment, which seems to have the greatest potential for negative social effects in that it entirely separates students into different streams" (p. 321). He claimed that there was little support for the assertion that high achievers benefit from streaming and low achievers suffer. "It is surprising to see how unequivocally the research evidence refutes the assertion that streamed class assignment can increase student achievement in primary schools. There is a considerable quantity of good quality research on this topic, such that any impact of streaming on achievement would surely have been detected" (p. 307).

Slavin (1987) also presented a meta-analysis of 14 studies relating to the Joplin Plan, most of which relate to grouping for reading (Table 4.3). The median effect size was .45, which is quite remarkable in the streaming literature. Further, the effects were consistently high, and there were no studies in which one subgroup gained at the expense of another – either all ability levels gained more than their control peers or (as in one study) none of the ability levels gained. The average effect for high-streamed students was .46, for middle it was .43, and for low it was .42. Slavin noted that one critical feature of the Joplin Plan (which we note is often not present in streaming) is frequent, careful assessment of student performance levels and provision of materials appropriate to these levels regardless of students' Year levels. The adaptation of instructional pace and level to student needs is considerable, and there is more movement especially up (and occasionally down) the levels. Kulik and Kulik (1987) reported an average effect of .23 from 16 studies based on the Joplin Plan.

Table 4.3

Summary of 14 Joplin-Plan Studies Reviewed by Slavin (1987)

Study	Year levels	Duration (months)	Overall	Reading	Maths	Effect size		
						High	Medium	Low
Morgan & Stucker (1960)	3-6	8	.30	.30		.32		.94
Hillson et al. (1967)	1-3	8-40	.72	.33				
Russell (1946)	4-6	16	.00	.00				
Green & Riley (1963)	4-6	8	.36	.36				
Ingram (1960)	1-3	24	.55	.55				
Halliwell (1963)	1-3	8	.59	.53				
Carson & Thompson (1964)	4-6	8	.00	.00				
Anastasiow (1968)	4-6	8	.15	.15				
Hart (1959)	4-5	8	.89	.89				
Rothrock (1961)	4-5	8	.44	44				
Skapski (1960)	3	24	.57			.91	.48	.52
Hart (1962)	1-3	24	.46		.46			
Moorhouse (1964)	4-6	2.5	.63			.61	.74	.34
Kierstead (1963)	3-8	8	-.02			-.01	.08	-.14

Slavin (1990) conducted a similar meta-analysis for secondary schools, where streaming is often adopted as part of a whole-school approach. Across the 29 studies, the typical effect size was –.02 (Table 4.4). This near-zero finding was the case in schools where all subjects were streamed and in schools where only some subjects were streamed. The effects for high (.01), average (–.08), and low (–.02) achievers were not different from zero, and the average effect for both reading and maths was .01. Slavin concluded that comprehensive between-class streaming has little or no effect on the achievement of secondary students, at least as measured by standardised tests. Further, there was "little support for the proposition that high achievers gain from grouping whereas low achievers lose" (p. 486). Streaming is equally ineffective in all subjects (except that there may be a negative effect of ability grouping in social studies), and thus it appears that it "simply does not matter whom students sit next to in a secondary class" (p. 491). In "study after study, including randomised experiments of a quality rarely seen in educational research, [there is] no positive effect of ability grouping in any subject or at any grade level, even for the high achievers most widely assumed to benefit from grouping" (p. 491).

Gamoran (1987a, 1987b) was critical of Slavin's (1990) review because it did not distinguish between school or class organisation. "Grouping does not produce achievement: instruction does" (p. 341). Hence, the extent that the effects of grouping are mediated through teachers' instructional behaviour is critical and, he argued, the study of grouping alone provides little information of value. This may be the case; however, the overwhelming summation of these meta-analyses indicates that this is highly unlikely – there is too little variance in the overall close-to-zero findings. If teachers are using differential instruction, it does not appear to be having an effect on the achievement outcomes of the students.

Table 4.4

Summary of 29 High School Studies Reviewed by Slavin (1990)

Study	Year levels	Duration (months)	Overall	Reading	Maths	Effect size High	Effect size Medium	Effect size Low
Marascuillo & McSweeney (1972)	8-9	16	-.22			.14	37	43
Drews (1963)	9	8	-.03	-.11		.16	.01	-.01
Fick (1963)	7	8	-.01	-.01		.01	.00	.04
Peterson (1966)	7-8	8	-.04	.02	-.25	.05	-.44	-.06
Ford (1974)	9	8	.00		.00			
Bicak (1962)	8	2.5	-.25			-.39		-.10
Lowell (1960)	10	8	.00					
Billett (1928)	1	8	.04	.04		-.11	.03	.18
Platz (1965)	9	2.5	.22			.24	-.10	.22
Bailey (1968)	9	8	-.03		-.03	.18		-.24
Thompson (1974)	11	8	-.48			-.50	-.45	-.54
Barton (1964)	9	8	-.04	.06		.22	-.08	-.2
Willcutt (1969)	7	8	-.15		-.15			
Holy & Sutton (1930)	9	2.5	.28		.28			
Martin (1927)	7	8	10	.17	.13	.12	-.06	.23
Kerckhoff (1986)	5-10	40	.03	.02	.03			
Fogelman et al. (1978)	6-10	32	.03	.02	.03			
Borg (1965)	6-12	32	.00					
Ferri (1971)	5-6	16	.00					
Breidenstine (1936)	7-9	8	-.19					
Purdom (1929)	9	2.5	.01	-.02	.00	-.02	-.08	.07
Postlethwaite et al. (1977)	5-7	16	.00					
Bachman (1968)	7	8	.00					
Kline (1964)	9-12	32	.01	-.05	.01	.03	.00	-.02
Stoakes (1964)	7	8	.00					
Martin (1959)	6-8	16	.00					
Chiotti (1961)	9	8	.18		.18	.14	-.18	-.05
Fowlkes (1931)	7	2.5	-.20	-.04	-.17	-.45	-.18	-.05
Cochran (1961)	8	0	.00					

Noland and Taylor (1986) calculated 720 effect sizes from 50 studies of streaming across primary and secondary schools. They selected studies published since 1967 that were conducted in the United States, and thst used comparisons between homogeneous and heterogeneous groups in Kindergarten (Year 1) to Year 12 classes. Unlike the earlier reviewed meta-analyses by the Kuliks and Slavin, they did not average the effect sizes across the study; hence there is more detail on the studies and more control of the various moderators. The average effect size across the 720 effects and across all outcomes was –.08 (the average effect size using one average effect size per study (as the Kuliks and Slavin had done) was .00). This zero to slightly negative effect occurred for both achievement (.01) and affective outcomes (–.15) across all Year levels. Table 4.5 presents further breakdowns of effect sizes for all outcomes, for cognitive, and for affective outcomes.

For the higher- (.16) and lower-streamed students (.18), the cognitive outcome scores were higher than for the middle-streamed students (–.45); the latter group being negatively affected by streaming. The affective outcomes were negative for all groups of streamed students: high (–.09), average (–.15), and low (–.35). The only exception to this zero to slightly negative pattern was for reading scores, where the effect size for cognitive (.06) and affective (.16) outcomes were slightly higher. Students in schools in lower socio-economic communities who were streamed scored higher on cognitive (.31) and affective (.24) outcomes than students in higher socio-economic communities (–.02 and –.17, respectively, for the two outcomes). The cognitive outcomes for the female students (–.31) were more than twice as negative as those for the male students (–.12). Noland and Taylor reported that researchers who appeared to favour streaming tended to find evidence for increased achievement (number of effects = 349, effect size = .08), whereas researchers who appeared to oppose streaming found heterogeneous grouping more beneficial to students (number of effects = 346, effect size = –.24). This is probably because the researchers were influenced by the findings of their own studies and framed their review in a manner that led to the conclusion – this bias should not necessarily be considered ill-intentioned.

Noland and Taylor's (1986) overall conclusion to the question 'Does ability grouping work?' was 'No'. Streaming, "while favoured by most teachers and entrenched in the public schools of the United States, does not (except in some very specific circumstances) improve student achievement and has potentially serious negative self-concept consequences.... [Instead,] we ought to be seeking policies and programs which enhance educational outcomes and which promote fairness in educational processes. Ability grouping does neither" (pp. 29–30).

Table 4.5

Effect sizes for Studies Reviewed by Noland and Taylor (1986)

	All		Cognitive		Affective	
Attribute	n	Mean	n	Mean	n	Mean
All outcomes	720	-.08	272	.01	402	-.15
Attitude towards subject matter					15	.22
Attitude towards school					13	.03
Attitude towards others					41	.09
Global self-concept					56	-.04
Academic self-concept					133	-.30
Years 1-5	133	-.04	84	.00	49	-.11
Years 5-7	196	-.11	58	-.20	120	-.11
Middle school	24	.08	24	.08		
Junior High	298	-.08	87	.22	191	-.20
High School	69	-.15	19	-.34	42	-.05
Subject: Reading			83	.06	6	.16
Subject: Math			63	-.11	22	-.45
Subject: Language Arts			33	-.01	20	-.12
Subject: Social Studies			40	-.06	67	-.08

Table 4.6

Summary of 10 Studies Reviewed by Mosteller, Light, and Sachs (1996)

Study	Year levels	Subject	Duration (months)	Overall	Effect size		
					High	Medium	Low
Barton (1964)	9	English	8	.11	.29	.02	.05
Bicak (1962)	8	Science	4	-.33	-.55	-.33	-.16
Drews (1963)	8	English	8	-.04	-.18	.02	-.08
Fick (1962)	7	Core	8	.02	.25	.09	-.27
Ford (1974)	9	Math	8	.29	na	na	na
Lovell (1960)	10	Alg, Bio, Eng	8	.14	.23	.14	.04
Marascuilo & McSweeney (1972)	8,9	Social Studies	16	-.16	.02	-.20	-.30
Peterson (1966)	7,8	Lang, Hist, Math	8	-.10	.14	-.42	.02
Vakos (1969)	11	History	8	.09	.10	.08	.10
Wardrop et al. (1967)	3	Math	2	.28	-.01	.42	.42

Mosteller, Light, and Sachs (1996) had stricter selection criteria for choosing studies. They chose only studies that included a treatment and a control group and that had been designed as a randomised field trial – that is, the assignment of students to the treatment and control groups (streamed versus whole class grouping) must have been either randomised or a close approximation to randomisation. Thus, they excluded studies that 'matched' students, on the ground that this method does not guarantee initial equivalence of groups. Ten studies met these criteria (Table 4.6). The average effect size, weighted by sample size, was zero. The effects for the high, medium, and low groups are also close to zero. Overall, they concluded that there was "little evidence that skill grouping has a major impact, either positive or negative, on students' cognitive learning" (p. 812).

4.1.2 Effectiveness studies: Correlational studies

There have been many studies that have used regression designs whereby a sample of students, often within a few schools, are administered a number of questionnaires and achievement tests. There are also many studies that have used larger secondary databases. Perhaps the most cited recent study has been Argys, Rees, and Brewer (1996). They claimed that a major methodological problem with most studies on streaming was that they did not control for the quality of the teachers, particularly given that many studies have found that low-stream classes tend to be assigned the least qualified teachers and receive less than their share of educational resources. They used 3,405 Year 11 students from the NELS (1988) (see Ingels, 1988; Russo, 1988) database and compared their achievement relative to their performance when they were in Year 9. They partialled out the effects of differential teaching quality, and calculated selectivity correction terms to assess the effects of streaming on student achievement, controlling for the selection process. They found that prior achievement played a critical role in determining stream placement, although previous stream placement also contributed, as did socio-economic status; students from better backgrounds were more likely to be placed in upper-level streams. After holding socio-economic status constant, race and ethnic background were not good predictors. This should not be surprising in the United States, as socio-economic status and race are often highly correlated. They concluded that "streaming is an important determinant of student achievement", and that, "if all students in our sample were placed in heterogeneous classes, average scores in mathematics could be expected to decline by approximately 2 percent" (p. 640). Hence, they claimed that detracking (creating, or reverting to, unstreamed / untracked classes) is clearly not a costless solution. The effect size, however, was only .03, which hardly justifies the strength of this conclusion.

Argys, Rees, and Brewer (1996) relied on statistical significance to determine important findings, but this method is rather meaningless given the large numbers of students (where most effects are statistically significant). It would have been more informative to have estimated effect sizes. Jaeger and Hattie (1996) estimated the overall effect size in the Argys et al. study as .13; which is not inconsistent with the overall findings of the meta-analyses reported above.

One of the difficulties of secondary analyses of large databases, such as NELS, is that the researchers must use the same questions posed when the database was compiled, which may not be exactly relevant to the purpose of the secondary analysis. In the Brewer et al. (1995) study, they used a proxy question for streaming – 'which of the following best describes the achievement level of the Year 11 students in this class compared with the average Year 11 student in the school – higher achievement levels, average achievement levels, lower achievement levels, or widely differing achievement levels?' The use of this proxy creates a substantial risk of confounding the presence or absence of streaming with several plausible alternatives. As Jaeger and Hattie (1996) have pointed out, it is possible that responding teachers regarded most of their students as comparatively sharp or dull, with little correspondence to any objective basis for students' classification. It is thus not clear that teachers' responses to the question cited can be regarded as indicative of streaming. More to the point for Year 11 mathematics, students are substantially differentiated by mathematics course content. Some students will be enrolled in a second or third course in algebra, some in a first course in algebra, some in business mathematics, some in general mathematics, and some in remedial mathematics. Hence, there is little assurance that the differences in mathematics performance reported by Brewer et al. were due to the effects of streaming rather than to the effects of the mathematics curriculum.

Gamoran and Mare (1989) used 10,980 adolescents from the HSB database, with particular emphasis on mathematics achievement and high-school dropout. Because of the nature of the questions asked in the HSB, they could only survey differences between college and non-college streams. The researchers used various regression-based models with particular emphases on assessing the influences of prior individual, school-level conditions, student and school social demographic factors, school programme emphasis, race, gender, ethnicity, and socio-economic status. They found that the school-level factors had little effect on outcomes once student-level measures of prior achievement and demographic factors were taken into account. Not surprisingly, adjusting the differences in senior mathematics achievement for variation in junior achievement test scores reduced the gap between the streams by 73%. More interestingly, socio-economic status accounted for five percent of the variation, while the other demographic variables contributed negligible amounts. Our estimate of the effect size of streaming in this study is .20, and thus disagree with Gamoran and Mare's conclusion that the "net track effect on mathematics achievement is substantial" (p. 1172).

The differential effects of streaming on graduation, after partialling out achievement differences, were negligible. Thus, Gamoran and Mare (1989) concluded that all students would be more likely to graduate if they enrolled in the college stream. Further, low achievers were less likely to be assigned to the college-preparatory programme, and thus streaming reinforced initial differences among students assigned to college and non-college curricula. "Tracking widens the gap in achievement and in the probability of graduating between students of high- and low-SES backgrounds" (pp. 1176–1177). Further, because the assignment process favours blacks and females over non-blacks and males (the former are more likely to be assigned to college and the latter to non-college streams), current streaming practices produce less inequality than would occur if students were randomly assigned to streams. These positive selection effects, however, do not overcome the stream effects themselves, which favour the college stream.

These regression-based methods are aimed at controlling for pre-existing differences in student achievement, but this method cannot be as powerful as randomly assigning students to the various streams. In a rare study employing random assignment of students to stream, Mason, Schroeter, Combs, and Washington (1992) placed 34 average-achieving Year 9 students into high-stream pre-algebra classes with their high-achieving peers. Several of these average achieving students performed better than their high-achieving peers and "took substantially more advanced mathematics during high school" (p. 597). The high-achieving students suffered no decrease in computation or problem-solving achievement and scored higher in concepts than their cohort 'average' peer groups from previous years.

On the basis of a review of survey and ethnographic research, Gamoran and Berends (1987) concluded that, when appropriate controls for prior achievement are incorporated, most of streaming's influence on academic achievement disappears. They summarised data from 10 American data sets used in 16 studies (including NELS, Project Talent, Youth in Transition, and HSB). These data spanned 29 years, although no time trends were apparent. Gamoran and Berends noted that, by reducing the pace and complexity of classroom work, teachers believe they are gearing instruction to the ability levels of the students. It may also assist in controlling students' behaviour: "Teachers used structured written work as a device to quiet a class or keep it calm" (Metz, 1978, p.103), particularly in low-streamed classes. Low-stream students appear to prefer such a pace as it is "less taxing and creates a sense of routine. Moreover, low-track students preferred written work because it was more private. In the oral instructional engagements in the higher tracks, mistakes were more visible" (p. 423). Such slower-paced instruction means that important parts of the curriculum may be introduced later for low-stream students, which can have an accumulative retrogression effect on these students' later chances at attaining educational desirability such as access to more challenging upper-school courses and university entrance. Such pacing also destines the students to remain in the lower streams, as they are now even further behind their middle-streamed peers.

Hoffer (1992) used the Longitudinal Study of American Youth database, which is from a four-year panel study of middle and high school science and maths education, started in 1987. A sample of 5,945 Year 8 and 9 students from over 100 schools was included. As expected, the high-stream students started far ahead of low-stream students at the start of the year, but they also showed greater gains within the year, hence moving even further ahead. The science effect sizes for the low-stream versus ungrouped classes were –.40 for Year 8 and –.17 for Year 9 students and .08 for the high-stream versus ungrouped classes for students in both Years. The maths effect sizes were –.36 and –.32 for the low group and .26 and .18 for the high group (Year 8 and Year 9, respectively). Hence, differential effects were found in both subjects, but more so in maths than in science. Hoffer noted, however, that there were more students in the high stream than the low stream, concluding that the "net effects of grouping turns out to be about zero…. Ability grouping in seventh- and eighth-grade mathematics and science is clearly not an optimal arrangement compared with the non-grouped alternative, for low-group students are significant losers" (p. 221). After testing many other models for moderators, they concluded that they could find no conditions under which grouping benefits all students (or at least helps some and does not hurt any).

In one of the few studies to use HLM to ascertain the effects at both the student and the class level, Bode (1993) used performance in mathematics for 1,319 Year 9 students in 79 classes from 61 schools. There were no advantages to being in either a streamed or unstreamed class. One of the best predictors was high level of effort (rather than prior achievement), and the amount of time spent in small-group instruction in heterogeneous rather than homogeneous classes. There were no effects relating to class, teacher, or instructional programmes.

4.1.3 The effects on higher-ability students

At best, streaming benefits the most advantaged students, although we note that these effects are small. Those in the higher streams have the best defined and most carefully sequenced programmes, partly because university-bound entrance criteria are clearer than vocationally-bound entrance criteria; the curriculum requirements are more shared and understood by teacher and students, and the students and their parents are more effective in seeking advantages. Specific classroom instructional interventions, and probably the quality of teaching, however, are more powerful than the streaming of students. For example, enrichment had less impact when students were placed in streamed classes. In a summary of the 593 effects across the various meta-analyses based on gifted programmes (Hattie, 1992), the average effect size for achievement was .35 (se=.16, n=490) and for attitude .14 (se=.15, n = 103) (Goldring, 1990; Kulik & Kulik, 1984; Rogers, 1991; Vaughn, 1990; Vaughn, Feldhusen, & Asher, 1991). The average effect size over all 593 effects was .38 (se=.17). The effects for streaming of gifted students was .30 (n =191), accelerated instruction was .22 (n =176), and enrichment within regular classes was .26 (n = 30). Enrichment activities within regular classrooms had much greater effects than streaming, with particularly larger effects on affective outcome measures. Enrichment programmes in which the students mastered successively more complex and challenging ideas had the largest effect, and programmes that involved the students in a broader investigation of the regular curriculum had slightly less impact. Enrichment instruction that occurred between six and 10 hours a week was most beneficial (1.23), and programmes lasting three months in duration also had large effects (.86).

It is important to note that the positive effects of streaming for the higher-ability groups are very likely to be confounded with the effects of gifted education programmes. It seems that all meta-analyses have included programmes for gifted students as examples of streaming. There can be marked differences between students who are streamed into high-stream classes and students who are identified as 'gifted'. The former typically receive a faster pace of instruction and more challenging tasks within the same curriculum frameworks as medium- and low-ability students, whereas the latter often have different curricula. Note, for example, the synthesis of 13 meta-analyses of gifted students reported by Rogers (1991). She noted "substantial academic gains" for gifted students enrolled full-time in special programmes for the gifted and talented (Kulik, 1985; Kulik & Kulik, 1982a, 1982b, 1984, 1990; Vaughn, 1990). Goldring (1990) reported effect sizes of .12 for gifted students after one year in special gifted programmes, and .47 for students who participated for more than one year. The greatest advantage was in science and social science tests, and the smallest advantage was gained in reading and writing. Gifted students in special classes exhibited better attitudes towards school, but they had more negative attitudes towards their peers than did gifted students in regular classrooms. Wallace (1989) found an effect size of .57 for gifted students who received enrichment in special classes compared with the achievement of gifted students in non-enriched classes. Again, the effects were highest in science and mathematics. It is noteworthy, that Wallace reported that heterogeneous grouping of gifted students of varying abilities had the largest effect size (.75), and for the model where gifted students are pulled out of their regular classes, the effect size was nearly as large (.68). Enrichment had less impact when students were grouped homogeneously (.23) or attended a special school (.10).

Kulik and Kulik (1992) found that multi-tracked classes, which entail only minor adjustment of course content for ability groups, usually have little or no effect on student achievement. Programmes that entail more substantial adjustment of the curriculum to ability, such as cross-year and within-class programmes, produce clear positive effects. Programmes of enrichment and acceleration, which usually involve the greatest amount of curricular adjustment, have the largest effects on student learning. Kulik and Kulik (1993) noted that the strongest benefits for gifted students were in "programs in which there was a great deal of adjustment of curriculum" and when there was "acceleration of instruction" (pp. 2–3).

4.1.4 The effects on lower-ability students

Kerckhoff (1986) in a study of 4,000 students in British schools used regression analyses and concluded that students in low-streamed classes lose ground, and that those in high-stream classes increase their average performance level beyond that exhibited by comparable students in ungrouped class settings. Hoffer (1992) found similar findings in United States schools, in which upper-stream classes had gains but lower-stream classes had decreases in scores compared to ungrouped classes. In science, for example, the effect size comparing students in ungrouped classes to those in low-streamed classes was .41 (in favour of non-streamed classes, after holding constant socio-economic status, school size, etc). The claim is that top streams accelerate achievement and low streams reduce achievement (Alexander, Cook, & McDill, 1978; Dar & Resh, 1986; Gamoran & Berends, 1987; Gamoran & Mare, 1989; Oakes, 1982; Sorenson & Hallinan, 1986).

The results from the regression and qualitative designs point to the zero or negative effects of streaming for lower-streamed students. The authors of most meta-analyses often highlight the negative effects. The data, however, are less supportive of this conclusion. The overall effect size for the five meta-analyses that report the effects for these students is .08 (.18, .14, .08, .02, -.02) – hardly convincing of a negative effect. More concerning is that these effects are probably lowered because of the lower quality of instruction that occurs in lower-streamed versus heterogeneous classrooms. Lower-streamed classes typically receive lower quality of instruction than do students in higher-streamed classes (Evertson, 1982; Gamoran, 1989; Oakes, 1985; Trimble & Sinclair, 1987). It may be, as Slavin (1990) has argued, that teachers cover less material in lower-stream classes, and students may be more off task (which hardly seems surprising).

It is likely that the very instructional advantages that accrue for high-stream students can be realised similarly for low-stream students. For example, Levin (1996) has extensively documented the advantages of accelerated learning for low-stream students. The present low-stream system is an intervention that assumes that low-stream students will not be able to maintain a normal instructional pace without further prerequisite knowledge and learning skills. The students "are placed into less demanding instructional settings – either by being pulled out of their regular classrooms or by adapting the regular classroom to their 'needs' – to provide remedial or compensatory educational services. This approach appears to be both rational and compassionate, but it has exactly the opposite consequences" (p. 333). Levin identified many of the factors that were successful in high-stream classes and implemented these for all students. These factors include a pedagogy not of remediation but acceleration that capitalises on the child's strengths, strategies for building on parental involvement in the school and in the home, assessment systems that monitor both school and student progress, a clear timetable for enhancing students' competencies for reaching specific goals, building the school's capacities to establish a unity of purpose and to make responsible decisions, enlisting higher standards for all students, and building on the child's experiences, interests, motivations, culture, and observed abilities. There are many studies supporting the positive effects of these implementations of accelerated learning (Chasin & Levin, 1995; Knight & Stallings, 1995; McCarthy & Still, 1993).

4.1.5 Teaching and learning across the streams

Perhaps the most influential in-depth study of teaching and learning in streamed classes is Oakes' (1985) *Keeping track: How schools structure inequality*. Her study was based on an intensive qualitative analysis of 25 junior and senior high schools, that was part of the Study of Schooling (Goodlad, 1984). The major finding was that many low-stream classes are deadening, non-educational environments. Oakes (1992) concluded that "the best evidence suggests that, in most cases, streaming fails to foster the outcomes schools value" (p. 13). Streaming fosters friendship networks linked to students' group membership, and these peer groups may contribute to 'polarised' stream-related attitudes among high school students, with high-stream students becoming more enthusiastic and low-stream students more alienated (Oakes, Gamoran, & Page, 1992). In subsequent evaluations, Oakes (1993) commented that streaming limits "students' schooling opportunities, achievements, and life chances. Students not in the highest tracks have fewer intellectual challenges, less engaging and supportive classrooms, and fewer well-trained teachers" (p. 27). Shanker (1993), President of the American Federation of Teachers, in a commentary of Oakes' research, was more earthy: "Kids in these [lower] tracks often get little worthwhile work to do; they spend a lot of time filling in the blanks in workbooks or ditto sheets. And because we expect almost nothing of them, they learn very little" (p. 34). In a similar qualitative design, Page (1991) provided a detailed account of daily activities of eight low-stream classes and found that teachers and students came to understandings about how to not push each other too hard so that they could cope, that low streams were used as 'holding tanks' for students with the most severe behaviour problems, and that teachers focused on remediation through dull, repetitious seatwork (seel also, Camarena, 1990; Gamoran, 1993).

The nature of instruction is different in the various streams. Gamoran, Nystrand, Berends, and LePore (1995) conducted a two-year study of instructional methods in 92 high-, regular-, and low-streamed classes in 25 secondary schools. They found fewer differences in the quality of instructional discourse than anticipated, although discussion time was more evident in the high-stream classes. High-stream classes had more coherence, in that the teachers framed lessons and activities in terms of previous lessons and activities. It was less the interactive style than the content of the interactions that favoured higher-stream over regular- and lower-stream classes. In the high-stream classes, there was more instruction relating to the subject matter. In low-stream classes, instruction was more often fragmented, emphasising isolated bits of information instead of sustained inquiry (Page, 1991). In studies by Gamoran (Gamoran, 1989; Nystrand & Gamoran, 1988), students in low-stream Year 9 and Year 10 English classes answered true-false, multiple-choice, and fill-in-the-blank questions four to five as frequently as did their high-stream counterparts. In low-stream classes, teachers commented about twice as much about spelling, punctuation, and grammar and about half as much as about content, compared to teachers' responses to high-stream students' papers.

Urdan, Midgley, and Wood (1995) worked collaboratively for three years with staff of a middle school who wanted to critically examine their policies on streaming. They found that "tracking affects the way teachers think about instruction" (p. 25) and made the entire school schedule less flexible. Gamoran (see also Lockwood, 1996) found that teachers in low-stream classes typically believed that teaching academic material is not the main goal for their classes, rather the goal is to keep their students well-behaved, interested, and achieving at the C-level. Only two of the 25 low-ability classes that he studied over a two-year period had teachers who held higher academic standards and required their lower-streamed students to work hard. Spear (1994) found that teachers who wished to retain streaming were more subject-centred, and those who wished to eliminate streaming were more student-centred.

It is worth noting that the major reason why teachers attest to the value of streaming is that it is more likely that they can provide instruction closely suited to the readiness and needs of different students. The most typical effect of streaming on teaching, however, is that students in lower streams receive more 'slowly' paced instruction rather than varied instructional opportunities. The same curricula offered to higher streams is depleted of task complexity when offered to lower streams, which can lead to maintaining students in their streams, and encouraging more experienced teachers to seek the challenges of the higher streams.

More experienced teachers do prefer the challenges of teaching to higher streams. Teachers of the higher streams report devoting more time and energy to preparation, claim greater satisfaction, are more responsive to challenging feedback and questions, and are more likely to have greater promotional opportunities because of these factors. Teachers' perceptions may also contribute to a polarisation in that they view high-stream students positively and low-stream students negatively (Ball, 1981; Brown & Goren, 1993; Finley, 1984; Gamoran & Berends, 1987). Instead of a cognitively, developmentally appropriate curriculum that challenges lower-stream students, these students are subjected to less challenging curricula and instruction (Bigelow, 1993; Dawson, 1987; Gamoran, 1986, 1992). Students in the lower streams are seen by teachers as 'thick' and 'unresponsive' (Finley, 1984; Schwartz, 1981), and teachers devote more time to discipline than content (Coley, 1991; Johnston & Markle, 1983; Lindle, 1994; Lockwood, 1990; Oakes, 1981). Such stereotypic conceptions can influence the teachers' expectations of the students' performance. For example, Tuckman and Bierman (1971) randomly selected high school students from a lower-ability class and assigned them to the next-highest class. At the end of the year, 54% of these randomly selected students remained at the higher level, but only one percent of the comparison students in the promoted students' original class were recommended for placement into the higher group in the next year. Streaming placements usually remain stable, partly because early assignments shape students' later school experiences (Oakes, 1992), and the attitudinal differences are likely to be reinforced and perpetuated as students initiate and maintain within-stream friendships (Hallinan & Williams, 1989).

Teachers of lower-stream classes are themselves often lower in ability and experience, set lower academic standards for their students compared with standards set by teachers in higher-stream classes, and have less access to class materials or science laboratories (Ball, 1981; Dornbusch, 1994;

Finley, 1984; Gamoran & Berends, 1987; Oakes, 1985; Rosenbaum, 1976). Oakes et al. (1990) found that "nearly all types of schools place their least-qualified teachers in low-ability classes and their most-qualified teaches in high-ability classes" (p. 65). The following table, from Oakes et al., illustrates the distribution of teachers by qualifications. High-streamed students get the best trained teachers, and low-streamed students in the most advantaged schools are likely to have better qualified teachers than high-streamed students in the least-advantaged schools. Thus there is a double disadvantage for low-streamed students in low-income, predominantly minority schools.

Table 4.7

Summary of Secondary Teachers' Qualifications in Low- and High-streamed Classes in Schools of Different Types (Oakes et al., 1990, p.66)

	Low Streamed Classes		High Stream Classes	
Teacher Qualifications	Low SES Minority Urban	High SES White Suburban	Low SES Minority Urban	High SES White Suburban
Certified in maths/science	39	82	73	84
Bachelor's in science/maths	38	68	46	78
Masters in science/maths	8	32	10	48
National Science Teach Assn qualified	11	36	5	47
National Council Teach Maths qualified	23	26	4	16
Computer coursework	41	61	69	62

Finley (1984) noted that teachers competed against each other to obtain the higher-streamed classes – higher-stream teachers "struggled through informal processes to maintain a monopoly over their jealously guarded classes" (p. 239). Gamoran and Berends (1987) reported that once assigned to the high streams, teachers appear to put more time and energy into their teaching. They spend more time preparing for classes (Rosenbaum, 1976), and they are more ready to respond to the more challenging questions posed by high-stream students (Metz, 1978). Oakes (1985) concluded that upper-streamed teachers tend to be more enthusiastic, to vary their method of presentation, and to use more constructive criticism than teachers in lower-streamed classes. Similarly, Schwartz (1981) found that when high-streamed students gave incorrect answers, teachers pushed them to develop the correct answer. Low-streamed students whose answers were incorrect were ignored; the teacher simply went on to ask another student. Oakes et al. (1990) found that high-ability classes had less seatwork, fewer worksheets, and fewer tests or quizzes than low-streamed classes, and that teachers in low-streamed classes placed less emphasis on nearly the entire range of curriculum goals.

4.1.6 Ethnicity, social class, and gender

The differential achievements relating to racial and social class groups are among the most contentious issues related to the streaming debate. Oakes and Wells (1996) claimed that streaming exists to guarantee the unfair distribution of privilege, in that white and wealthy students benefit from access to high-status knowledge that low-income students and students of colour are denied. Oakes et al. (1990) analysed 1,200 public and private primary and secondary schools in the United States, and found that "minority classes were seven times more likely to be identified as low-ability than as high-ability" (p. 23). Given the stability of streaming, however, more often it perpetuates these distinctions. Those schools that stream often explain this ethnic subdivision by reference to past achievement, and thereby argue that streaming can maximise opportunities to alter this. If streaming leads to proportionally more lower socio-economic students or students from particular ethnic groups being placed in lower streams, then the use of streaming may serve to increase divisions along class, race, and ethnic lines (Haller & Davis, 1980; Rosenbaum, 1980). In his survey of streaming policy in California and Massachusetts, Loveless (1999b) concluded that there are massive contradictions, in that detracking is taking place in low-achievement schools, in poor schools, and in urban areas; whereas suburban schools, schools in wealthy communities, and high-achieving schools are staying with streaming – indeed embracing it. "This runs counter to the notion of elites imposing a counterproductive policy on society's downtrodden. If tracking is bad policy, society's elites are irrationally reserving it for their own children" (Loveless, 1999b, p. 154). Braddock (1990) found that schools with more than 20% of their rolls from minority groups are more likely to stream than those with fewer minority students.

Principals generally express support for the statement that streaming is beneficial for minority students, although they also agree that it often results in racially or ethnically identifiable streams (Crosby & Owens, 1991). Gamoran and Mare (1989) found that African-American students had a 10% advantage over white students in being assigned to the high streams, providing some evidence to counter the claims that streaming is necessarily disadvantageous to minority students (see also Oakes et al., 1992). Oakes et al. (1992) found that Asian students were more likely to be assigned to advanced courses than were Hispanic students with whom their test scores were equivalent. A disproportionate number of low socio-economic status and disadvantaged minority students occupy the lower streams and non-college streams (National Center for Educational Statistics, 1985; Oakes et al., 1992; Persell, 1977; Van Fossen, Jones, & Spade, 1987). Noland and Taylor (1986) noted an interaction effect between streaming and the socio-economic areas from which the students came. Students in schools in lower socio-economic communities who were streamed scored higher on cognitive (.31) and affective (.24) outcomes compared to students in higher socio-economic communities (–.02 and –.17, respectively).

Students of average ability from advantaged families are more likely to be assigned to higher streams because of actions by their parents, who are often effective managers of their children's schooling (Alexander et al., 1978; Baker & Stevenson, 1986; Dornbusch, 1994; Lareau, 1987; Useem, 1991, 1992). Further, schools with a larger proportion of minority and lower socio-economic students are less likely to have sufficient higher-level courses, which affects the probabilities of students entering higher classes. Moreover, the higher-stream programmes in these schools are often less rigorous than higher-stream classes in schools with fewer minorities and higher socio-economic students (Oakes et al., 1992).

4.1.7 Detracking

There is a rich, recent literature encouraging schools to detrack (Bates, 1992; Lindle, 1994; Wheelock, 1994), although it is rarely accompanied by evidence of enhanced learning. There are many major United States reforms that have identified detracking as pivotal to reforming education (e.g., Goodlad, 1989). Further, there is an increasing number of successful United States legal challenges that have led to compulsory detracking (Welner & Oakes, 1996). As an indication of this trend, Loveless (1993) sampled 373 Californian middle-school principals and noted an increase in the proportion of English classes that were untracked, from 27% in 1986 to 48% in 1991.

If streaming has negative or at best only minimal gains, then it would be expected that schools that detrack should show also show minimal gains. Oakes and Wells (1998) followed 10 schools as they detracked: creating new schedules, reorganising teachers into teams, providing all students access to honours programmes, instituting integrated curricula, and creating opportunities for students to get extra academic support. Such changes were resisted by teachers and parents with fixed (i.e. intelligence is innate) rather than changeable conceptions of intelligence (50% of white Americans regard African-Americans and Latinos to be less intelligent that whites; Fulwood, 1991). Parents of identified gifted students severely criticised the schools for not offering separate enrichment classes, and the major reason for this concern was not the quality of the curricula, but that their children were no longer singled out and treated differently. Where the detracking was stopped, it was access to privilege, not the absence of challenge or quality that helped stop it. The battle was over which students (regardless of ability) would have access to which curricula and which teacher: "At risk for the families of high-track students is the entire system of meritocracy on which they base their privileged positions in society. As this system begins to crack, these parents often employ tactics that make reform politically impossible" (p. 6). Roe and Radebaugh (1993) examined one school as it eliminated streaming and found that teachers reported positive social benefits, positive behavioural implications, and less parental competition. The teachers also felt that detracking had academic benefits due to the social nature of learning and the strong influence of the adolescent's peer group (see also Bellanca & Swartz, 1993). Loveless (1999b) concluded that "we do not know very much about untracked schools. They have not existed in sufficient numbers to find out, as tracking reformers claim, that low-achieving, poor, or urban students benefit from heterogeneous grouping" (pp. 154–155).

4.1.8 Conclusions relating to Streaming

The overwhelming message is that streaming has minimal effects on learning outcomes. Across the meta-analyses of streaming (excluding the Joplin Plan) the average effect size is .05 (se=.03, n=261 studies, 784 effects; Table 4.8). When weighted by the number of studies included in the meta-analyses, the average was –.03. This overall effect of streaming must be interpreted relative to the typical effect size of .40 outlined in Chapter 1, and is therefore among the lowest of effect sizes across various interventions. There is, therefore, little confidence to consider that the overall effects of streaming have any noticeable effects on achievement. Across the five meta-analyses that provided sufficient information, it was possible to identify 75 unique studies, based on 45,000 students, typically conducted over 1.5 years. The average across these studies is .07, with a standard error of .03. The overall effects on maths and reading were similarly low (reading = .00, maths = .02), the effects on self-concept were close to zero, and attitudes towards subject matter slightly higher (.10).

Table 4.8

Summary of Achievement Related Effect Sizes in 7 Meta-analyses

Meta-analysis	Type	Level	No. studies	No. effects	Effect size	SD	SE
Kulik & Kulik (1987)	XYZ	Secondary	52	52	.10	.32	.05
Kulik & Kulik (1984)	XYZ	Primary	31	31	.19	.32	.03
Kulik & Kulik (1985)	XYZ	Primary/Secondary	85	85	.15	na	na
Slavin (1987)	XYZ	Primary	14	17	.00	.15	.04
Slavin (1987)	Joplin	Primary	14	14	.45	.29	.04
Slavin (1990)	XYZ	Secondary	29	29	-.02	.15	.03
Noland (1986)	XYZ	Primary/Secondary	50	570	-.08	.64	.03

The overall effects for the three major ability levels across the studies were .14 for high-streamed, –.03 for middle-streamed, and .09 for low-streamed students (Table 4.9). Table 4.10 depicts a stem and leaf diagram of the effect sizes for each level of schooling and for all studies. (hence, for primary schools there was an effect of -.31, -.34, -.26, -.10, -.13 etc.). With the exception of five studies at the primary level, the effects are most consistently centred around zero (to slightly negative). Much of the variability in the primary school is related to whether programmes for gifted students were included as 'high-ability' streams. Overall, the effects of gifted programmes are closer to .30, and when these special gifted programmes are removed, the average effect size for high-ability groups reduced to .02. For example, in Kulik and Kulik's (1984) meta-analysis, the mean effect size for programmes designed specifically for gifted and talented students was .49, whereas for high-ability streams for more representative populations, the mean effect size was .07. Similarly, in Kulik and Kulik's later (1985) meta-analysis, the effect size for gifted programmes (n=25 studies) was .33, and for XYZ programmes was .12 for high-ability streams. Similarly, Kulik and Kulik (1987) found

25 studies that investigated the effects of placing gifted students into special classrooms, and the average effect was .33, which was much higher than the effect of .12 from high-streamed classes.

Table 4.9

Summary of Effect sizes Relating to Low, Middle and High-ability Streams in 7 Meta-analyses

Meta-analysis	Effect size		
	Low	Middle	High
Kulik and Kulik (1987)			.33
Kulik and Kulik (1984)			.49
Kulik and Kulik (1985)	.14		.33
Slavin (1987	.02	.02	.03
Slavin (1990)	-.02	-.08	.01
Noland and Taylor (1986)	.18		-.16
Mosteller et al. (1996)	.08	-.04	-.06

Table 4.10

Stem and Leaf Display of Effect Sizes for Primary, Intermediate, and Secondary School

Effect size	Primary	Intermediate	Secondary	All
-.4			8	8
-.3	14	3		134
-.2	6		0	06
-.1	03		03569	003569
-.0	3468	4		334468
.0	000004555	00233	0000119	000000000001123345559
.1	5		0148	01458
.2	1378		289	1237889
.3	228			228
.4				
.5				
.6	018			018
.7	01			01
Count	31	7	21	59

The studies based on more rigorous research designs tended to produce effects much closer to zero, and the longer the streaming was implemented, then the lower the effects. This latter finding has been noted for many educational interventions, indicating that gains in achievement are likely to dissipate over time, particularly as teachers begin to accept the innovation and are then less attentive to changing their instruction to adapt to the intervention (Hattie, 1992). It must be noted that the data on which these meta-analyses are based are quite old. The median decade was the 1960s (and 13 of the 75 studies were published before 1940). The more recent Noland and Taylor (1986) meta-analysis did not provide the references of their articles, but it is likely that the articles used are somewhat more recent. Overall, the meta-analyses are based on dated literature. It seems ironic that Kulik and Kulik (1993) criticised Oakes because her conclusions are "based on her own selective and idiosyncratic review of older summaries of the literature" (p. 3). The more recent studies often have been based on longitudinal databases, and these too are more than 10 years old. The most recent summary of the streaming literature (Loveless, 1999b), which was more concerned with the policy debates relating to the 'tracking war', includes many more recent studies. Excluding the references relating to the history of streaming and the news release articles, the average year of Loveless's references is 1987, with the majority from 1980 to 1997. His review includes no effectiveness studies not cited in this review. There is much merit in the conclusion by Mosteller, Light, and Sachs (1996) that the main contribution from these studies of streaming is "a sharpened awareness of the limited amount of rigorous investigation that has been done" (p. 824). We would add that the studies of streaming are dated, the method appears entrenched in many schools, and the overall effects are consistently close to zero.

There are two major types of outcomes in the streaming debate: achievement and equity outcomes. The effects for both types of outcomes are close to zero. Such findings make it difficult to claim that streaming perpetuates social class and ethnic divisions, and provides greater learning opportunities for those students who are already socially advantaged. At best, the effects are small in favour of homogeneous streams, but it would be difficult on the basis of the studies presented, based on many hundreds of studies, to feel confident that schools were undertaking a worthwhile activity by considering or using streaming. The effects of many teaching-related, instructional innovations substantially overwhelm the effects of streaming. The trade-off is not between closing the gap between low-ability and high-ability students versus raising overall student test scores – it is between policy makers attending to classroom organisation practices versus improving what happens once the classroom door is closed. Whether a school streams or not appears less consequential than whether it attends to the nature and quality of instruction in the classroom, whatever the within-class variability in achievement. It is almost certain that there are conditions of learning (such as specific and challenging goals, the presence of feedback, and structure in the activities) that are far more powerful. It is likely that procedures other than streaming can optimise the advantages of peer influences in learning, such as reciprocal teaching and scaffolding. It is more likely that good teaching is more powerful a factor and can be independent of the class configuration.

There is, however, much evidence that streaming for low-achievers often means a slower pace of instruction rather than different instruction; a reduced content of the curriculum rather than a different curriculum; and a lower quality, less-experienced teacher rather than enhanced teacher quality. Note, however, that despite this evidence, the effects of streaming on students in lower-streamed classes is still close to zero, and these effects are not that different from those in higher-streamed classes. There is much evidence that the learning goals for low-stream students is more related to social and behavioural goals, whereas the learning goals for high-stream students is more related to achievement and understanding relating to future opportunities to learn. As Loveless (1999a) noted, low streams often emphasise good behaviour and menial skills, while high streams prepare for college. These differences in learning environments may be most powerful in depressing the academic achievement of poor and minority students. There is evidence that, when students, regardless of ability, are provided with good teaching, specific curriculum opportunities, reduced heterogeneity within specific subjects (and not across all subjects for the whole day), the effects of learning are greater. Achievement gains are unlikely unless there are changes in curriculum and instruction to accompany streaming (or detracking). "Changing one's teaching is not like changing one's socks" (Cohen & Ball, 1990, p.334). Loveless (1999b) claimed that "each teacher's opinions on educational issues, including streaming, interlock to create a personal philosophy of schooling. To achieve widespread detracking, sceptical teachers must be persuaded to reject personal educational beliefs in favor of a radically different system of values" (p. 133). Rather than streaming (or detracking), a far better investment would be to improve the level of teaching proficiency.

If it is the case, as the evidence suggests, that low-stream students are more likely to miss out on careful curricula specification, higher quality teaching, and high expectations, then there is compelling evidence that streaming, at least for low-stream students, could suppress potential learning gains. The effects of accelerated learning for example, point to the potential interventions that can have a positive influence – particularly on these low-streamed students. Again, the importance of teachers and teaching is indicated, and not the compositional effect of grouping. Grouping merely changes the probabilities of learning outcomes accruing and it seems that these opportunities, whatever they are, are not being taken when achievement homogeneity is reduced. A major cost from deciding to stream, therefore, is the false assumption that something has been done that can benefit the students merely by a grouping composition effect. It is important to note, that if a school decides to remove streaming, there is no guarantee that there will be enhanced learning outcomes. Recall that the effect sizes are small in favour of streaming, so detracking is unlikely to make dramatic differences.

The importance of curriculum relevance has been highlighted in this review of streaming, and such relevance is not necessarily a consequence of more homogeneity within classes. Slavin (Slavin, 1987) concluded that "for ability grouping to be effective at the primary level, it must create true homogeneity on the specific skill being taught and instruction must be closely tailored to students' levels of performance" (p. 323). Perhaps a best example of this is the Joplin Plan (Floyd, 1954), which involves grouping students for reading across Year levels. Thus, all students in the school are timetabled for reading at the same time, and then groups are formed based on reading ability across Years. The average effect of the Joplin Plan, based on 14 studies, is .45. Hence, streaming can occur in a way that allows expectations to be raised and allows movement across reading levels, without disrupting the whole class structure or establishing classes that teachers do not want to teach. The other example of curriculum tailoring relates to gifted classes, where the effects are much greater than for high-streamed classes. It is the curriculum relevance and not just the reduction of homogeneity that seems to make the difference.

The effects of streaming on self-esteem also are near-zero overall, albeit slightly positive for lower-streamed and slightly negative for higher-streamed students. Higher-streamed students may become slightly less satisfied with themselves when taught with their intellectual peers; slower students may gain slightly in self-confidence when they are taught with slower learners. This is known as the 'little-fish-big-pond' effect. Marsh (1984a 1984b, 1987; Marsh & Parker, 1984) has extensively documented these effects, which he claims are particularly evident with selective schooling. Marsh hypothesised that students compare their own academic ability with the academic abilities of their peers and use this social comparison impression as one basis for forming their own academic self-concept. The effect occurs when equally able students have lower academic self-concepts if they compare themselves to more able students, and higher academic self-concepts if they compare themselves with less able students. For example, if average-ability students are in a high-ability class, then their academic abilities will be below the average of other students in the class, and this will lead to academic self-concepts that are below average. Conversely, if these students attended a low-ability class, then their abilities would be above average in that class and this would lead to academic self-concepts that are above average. Similarly, the academic self-concepts of below-average and above-average students will depend on their academic ability but also will vary with the type of class they attend. According to this model, academic self-concept will be correlated positively with individual achievement (brighter children will have higher academic self-concepts) but negatively related to class-average achievement (the same children will have lower academic self-concepts in a class where the average ability is high). Thus, equally able students tend to have lower academic self-concepts if they attend academically selective classes (or schools) than if they attend classes in which the average ability level is lower. Marsh (1987) argued that "for at least some children, the early formation of a self-image as a poor student may be more detrimental than the possible benefits of attending a high-ability school" (p. 292).

4.2 Class Size

The research on the effects of class size has been among the more voluminous in educational research, with consistent findings. The earliest empirical studies were published at the turn of this century (Rice, 1902), although there are earlier claims in the Talmud that the maximum size of bible classes was 25 students. Reducing class size is a catch-cry of many teachers, unions, and politicians. President Clinton recently announced a $US1.2 billion class-size reduction plan, although Republicans are likely to dramatically modify the proposal. The very first meta-analysis was related to class size (*Education Week*, 18(18), pp. 13–17). Glass and Smith (1978) synthesised 77 studies, leading to a total of 725 effect sizes. The average effect size was .09, but, more importantly, there was a non-linear effect. Reducing class sizes from 40 or more to 20 students led to close-to-zero increments in achievement, but when class sizes dropped to 15 students or lower, there were larger effects on achievement. Smith and Glass (1980) also synthesised 59 studies covering 371 effects relating to class size and non-achievement based outcomes such as interpersonal regard, quality of instruction, teacher attitude, and school climate. Table 4.11 presents the effect sizes with a class-size of 30 as the anchor point at the 50th percentile of effects for both the achievement and non-achievement outcomes (calculated from Glass & Smith, 1978). Thus, if a student from a class of 30 was placed into a class of 20 students, he or she would experience achievement benefits superior to 54% of students, and non-achievement benefits superior to 59% of students, who are taught in the class of 30.

Table 4.11

Summary of Effect Sizes for Various Reductions of Class Size on Achievement and Attitudes

Class Size		Effect size			
Large	Smaller	Achievement	Success Ratio	Attitude	Success Ratio
30	5	.84	38	.41	20
30	10	.26	25	.52	25
30	15	.13	13	.33	17
30	20	.04	4	.19	9
30	25	.00	0	.09	5
30	30	0	0	0	0

Smith and Glass (1980) concluded that achievement, attitude, teacher morale, and student satisfaction gains were appreciable in smaller classes – provided we recognise that 'small' means 10 to 15 students – with negligible gains from reducing class sizes as high as 40 to 20 students. This effect was greater in secondary than in primary schools, but the same across all subjects and across various ability levels. Hedges and Stock (1983) reanalysed Glass and Smith's (1979) set of studies using slightly more rigorous statistical methods, but found no differences to the earlier conclusions. A more telling criticism of the Glass and Smith meta-analysis was that the studies were of short duration, included one-on-one tutoring, and were in some cases non-class related (e.g., tennis). Slavin (1989a, 1989b) found only eight studies that met his inclusion criteria of lasting at least one year, involving a substantial reduction in class size, and involving random assignment or matching of students across larger and smaller classes. He concluded that substantial reductions in class size have a small positive effects on students (effect size = .13) and the effect was not cumulative and disappeared within a few years.

McGiverin, Gilman, and Tillitski (1989) conducted a meta-analysis of 10 studies of Indiana's Prime Time project. This project aimed to reduce class sizes to 14 in 24 Years 1 to 3 classes, and the students were followed over three years. They reported that Year 3 students who had been in smaller classes for two years had significantly higher achievement test scores than did students in larger classes, with an overall average effect size of .34 (see also Chase, Mueller, & Walden, 1986, December; Malloy & Gilman, 1989; Mueller, Chase, & Walden, 1988). It is difficult to credit this effect to class size, however, as the study had few controls. It is not clear that small classes were kept small for the entire day, and, while the average class size for the 'smaller' classes was set at 18, actual 'small' class sizes ranged from 18 to 31, and classes of 24 were considered small if there was a teacher aide to assist the teacher.

Project STAR (Student-Teacher Achievement Ratios) was motivated by Prime Time and began in Tennessee in 1985 (for a history of this innovation, see Ritter & Boruch, 1999). This project involved a random assignment of students to regular classes (22–24 students) or small classes (14–16 students) when the students entered Kindergarten (Year 1) and keeping this size for the next three years, when the students then moved into regular-sized Year 5 classes. Over 6,500 students in 331 classes and 79 schools participated in the programme (Finn & Achilles, 1990; Finn, Folger, & Cox, 1991; Word et al., 1990), and each school had one of the three types of classes studied in this experiment: a small class, a regular class with a teacher aide, and a regular class without a teacher aide. Finn (1998) demonstrated that smaller classes benefited students in Year 1 through Year 4 academically, and there were improvements in the students' expenditure of effort, initiative taking, and reduced disruptive and inattentive behaviour in comparison to students in larger classes. The effect sizes were all positive in favour of the small classes (Table 4.12), and greater for minority (close to double) compared to white students for all achievement areas; zero effects were found for motivation and self-concept. Finn and Achilles (1990) reported that the difference between minorities and whites in mastery rates on the Year 1 reading test was "reduced from 14.3% in regular classes to 4.1% in small classes" (p. 568). Across all comparisons, the smaller class advantage in Year 1 was approximately .15 to .18; for Year 2, .22 to .27; and for Years 3 and 4, .19 to .26. These overall effects (.15 to .27) are not that different from what would have been predicted on the basis of Glass's meta-analysis.

Table 4.12

A Comparison of Small and Regular Class Effect Sizes for Years 1 through 4 for White (W) and Minority (M) students in Project STAR

Scale	Group	1	2	3	4
Word Study Skills	W	.15	.16	.11	na
	M	.17	.32	.34	na
	All	*.15*	*.22*	*.20*	*na*
Reading	W	.15	.16	.11	.16
	M	.15	.35	.26	.35
	All	*.18*	*.22*	*.19*	*.25*
Mathematics	W	.17	.22	.12	.16
	M	.08	.31	.35	.30
	All	*.15*	*.27*	*.20*	*.23*
Motivation	W	.00	-.02	-.03	-.01
	M	.03	-.01	.07	.11
	All	*.01*	*.00*	*.01*	*.00*
Self-concept	W	.10	.07	.00	-.05
	M	.10	.05	.03	.04
	All	*.11*	*.07*	*.02*	*.02*

The effect sizes on the Stanford Achievement Test (SAT) and Tennessee Basic Skills First (BSF) Test for the small versus regular classes with and without teacher aides are presented in the Table 4.13. Overall, the effect sizes for small versus regular classes without a teacher aide are about .25 and with a teacher aide around .10; hence it may be that introducing teacher aides rather than reducing class size could be a more cost efficient way of producing positive benefits for students.

Hanushek (1999) is highly critical of the large attrition rate in the Project STAR data. He noted that slightly less than half of the original students in the experiment remained in the study until the end of the third grade (Year 4). Nye, Hedges, and Konstantopoulos (1999), in a five-year follow-up study, found that students who left the small classes had higher achievement than those who left the larger classes, suggesting that the observed differences are probably not due to attrition. Hanushek's more cogent criticism, which is as yet unanswered, is that although randomisation was used, unlike randomisation in other areas (e.g., medical science), it was not blind in this study. That is, teacher, parents, school officials, and students were obviously aware of the assignment to small or larger classes. Hence, the results could have been related to more resources going to the smaller classes, and other "more direct motivation and incentives of teachers and principals that could bias the results of the different treatment groups" (Hanushek, 1999, p. 153). He also noted the high likelihood of school effects influencing the conclusions. The students were randomly assigned but the schools had to volunteer to participate. There were 79 schools with Kindergarten (Year 1) experiments, and half (40) showed advantages for small classes and the other half for regular classes. His conclusion, therefore, is that "it is only slightly better than an even bet from the STAR data that the small class achievement will exceed that of the regular and the regular with aide classes in any of the sampled schools" (p. 159). Other econometric studies also show small effects from reducing class size, with effects clustering around .00 to .10 (Boozer & Rouse, 1995; Hanushek, Rivkin, & Taylor, 1996; Krueger, 1997).

Table 4.13

Summary of Effect Sizes for Reading and Maths for Small and Regular Classes With and Without Teacher Aides for the SAT (Stanford Achievement Test) and BSF (Basic Skills First Test)

	SAT Reading	BSF Reading	SAT Math	BSF Math
Small vs. regular size without aides	.30	.25	.32	.15
Small vs. regular size with aides	.14	.08	.10	.05

There does appear to be problems with the initial assignment of students to classes as, in both reading and mathematics, students in the smaller classes had significantly greater average achievement at the end of Year 1 (i.e. at the end of the first year of the programme, and reading and mathematics are not a large part of the Year 1 curriculum). These effects were maintained throughout the duration of the project. At best, therefore, smaller classes in Year 1 may make a difference, but there was no evidence of differential effects beyond this first year. It may be that this early advantage allowed students to learn the 'skills of the classroom' and learn how to cope in the learning environment. This is conjecture and would need much further support.

At Year 5, all of the students in the study returned to regular-sized classes. The Lasting Benefits Study followed many of these students, some through to Year 11 (Finn et al., 1991). The effect sizes in favour of those who had begun in smaller classes were primarily in the .10 to .15 range, indicating that there were positive effects of this early-age intervention, even when the small-class intervention was disbanded. The effects for student engagement in learning (initiative taking, lack of disruption, attentiveness) were greater in the smaller classes (effect size = .13) a year after the students returned to normal classes (Finn & Achilles, 1999). Wenglinsky (1997), in an analysis of production functions based on Project STAR, also reported positive effects for small classes at Year 5 but not at Year 9. Pate-Bain, Boyd-Zaharias, Cain, Word and Binkley (1997) followed a cohort through to Year 11, and concluded that students who had been in the smaller classes appeared to have maintained academic achievement advantages. This is somewhat misleading as the differences are 'appearances' and were not statistically significant differences. There is more compelling evidence that students from small classes were less likely to fail a Year level, less likely to be suspended, and more likely to take more advanced courses than their peers who were in regular and regular/aide classes.

A key issue in this class size debate is whether the teachers change the instruction when moving from larger to smaller classes, and whether this has an effect on learning outcomes. Glass, Cahen, Smith, and Filby (1982) reported evidence that, too often, the nature of the instruction did not change when classes were reduced from 40 to 20 students. A poor teacher with 30 students may remain a poor teacher with 20. Shapson, Wright, Eason, and Fitzgerald (1980), in an unusual study, randomly assigned teachers and students in Year 5 to one of four class sizes: 16, 23, 30, or 37 students. The students were randomly reassigned in Year 6 and, as well as achievement measures, ratings were made of teacher-student interactions and classroom behaviours. The teachers expressed more positive attitudes with the smaller classes and were more pleased with the ease of managing and teaching in the smaller-class setting. However, "the observation of classroom process variables revealed very few effects of class size. Class size did not affect the amount of time teachers spent talking about course content or classroom routines. Nor did it affect the choice of audience for teachers' verbal interactions; that is, when they changed class sizes, teachers did not alter the proportion of their time spent interacting with the whole class, with groups, or with individual pupils" (pp. 149–150). No differences were found in student satisfaction or affective measures, teacher activities, subject emphasis, classroom atmosphere, or the quality measures.

Bourke (1986), in an Australian study, noted the correlates of class size. Significant positive correlations included amount of noise tolerated, non-academic management, and teacher lecturing or explaining. Significant negative correlations were more numerous, including use of whole-class teaching, amount of homework assigned and graded, teacher probes after a question, teacher directly interacting with students, and positive teacher response to answers from students. Thus, in smaller classes, less time is spent on classroom management, and there are more and more protracted interactions with students. Most importantly, it is these teacher interventions that are critically different. Reducing class size increases the probability that these more positive teacher interventions will occur, but it does not guarantee them – and, too often, teachers do not change their habits of instruction when their class sizes are reduced.

In the Project STAR findings of classroom behaviours, Evertson and Folger (1989) reported that students in smaller classes initiated more contacts with the teacher for purposes of clarification, gave more answers to questions that were open to the whole class, more often contacted the teacher privately for help, were more on-task, and spent less time waiting for the next assignment (see also Achilles, Kiser-Kling, Owen, & Aust, 1994; Kiser-Kling, 1995). Overall, however, in the Project STAR analyses, teachers tended not to change their fundamental teaching strategies when given a small class (Finn & Achilles, 1999).

There is little evidence that instruction methods change when class size is reduced, although a large part of any improvement relating to smaller class sizes can be explained by improvements in student task engagement (Finn & Achilles, 1999). The most likely explanation for the increases in achievement in smaller classes (of about 15 students) is that smaller classes enhance engagement and reduce inattention and withdrawal behaviours (Finn, Pannozzo, & Voelkl, 1995; Lamborn, Brown, Mounts, & Steinberg, 1992; McFadden, Marsh, Price, & Hwang, 1992; Steele, 1992). In the smaller classes, it is more difficult for the teacher to ignore the non-participants. Betts and Shkolnik (1999) conducted an intensive analysis of the Longitudinal Study of American Youth, which includes surveys by teachers, principals, students, and parents about student and teacher behaviour in the classroom. Class size variations induced little change in how teachers allocated their time between new material, review, discipline, routine tasks, and testing. In smaller classes, teachers did not increase the proportion of time spent on new materials but allocated more time to reviewing activities. They found that smaller class sizes induced teachers to devote less time to group instruction and more time to individual instruction. Their evidence, however, demonstrates that teachers could make small classes "considerably more effective if they did not reduce group instruction to the extent that they do" (p. 209). Overall, they argued that, because teachers reallocate their time to such a small extent, this "may explain why it has been so hard in most past research to identify a positive and significant impact of class size reduction on student achievement" (p. 209).

Rice (1999) completed a similar analysis but used the NELS database. She found that class size does not appear to influence the instructional strategies in science classes, and there are very small effects in mathematics classes. The effects were more pronounced in classes of higher-ability students, suggesting that the teachers do not change their instructional practices for lower-achieving students no matter what the class size. Hargreaves, Galton, and Pell (Hargreaves, Galton, & Pell, 1998) also concluded from their British study of classroom interaction differences that the successful teachers of the larger classes had "difficulty in maximising the opportunities offered in the small class setting, largely because they were unfamiliar with having to cope with such small numbers" (p. 791).

Hanushek (1986b, 1997, 1999; Hanushek et al., 1996), has long maintained that there is little evidence to support the benefits on student learning of smaller classes. In a series of summaries of the literature, he found 78 separate estimates of class size effects based on value-added results. Of these, 12% were statistically significant and positive in favour of smaller classes, and eight percent were negative; 21% were not statistically significant but positive, and 26% were negative. Hence, there is "little reason to believe that smaller classes systematically lead to improvements in student achievement" (p. 148). When he added studies conducted within a single state in the United States, Hanushek concluded that "more studies actually suggest that small classes are harmful… [and that overall, there is] "no consistent or clear indication that overall class size reductions will lead to improved student performance" (p. 149). It is perhaps not surprising that the box-scores of positive and negative are so close, when it is appreciated that the effect sizes are also close to zero. The box-score method, however, is not preferable to ascertaining the effect sizes, as it loses too much information (e.g., an effect size of .99 is positive, and an effect size of –.01 is negative, but the box-score concludes that there are as many positive as negative conclusions!).

The Wisconsin Student Achievement Guarantee in Education (SAGE) programme was designed as a five-year project commencing in the 1996–97 school year. Schools were required to implement four interventions: reduced class size, opening from early in the morning until late in the evening, developing rigorous curricula, and creating a system of staff development and professional accountability. Molnar, Smith, Zahorik, Palmer, Halbach, and Ehrle (1999) reported on the first two-year evaluation. They began by noting that only reduced class size (under 15 students) was uniformly implemented across all programme schools. From a series of regressions and an HLM analysis, they concluded that the effect size from the first-year SAGE students for class size reductions was about .2, and higher for African-American students. Interestingly, they noted no differences between class sizes of 15 with one teacher and class sizes of 30 with two teachers, concluding that this "suggests that the benefits of reducing class size may be achievable without the attendant capital costs of building additional classrooms" (p. 177).

Brewer, Krop, Gill, and Reichardt (1999) estimated the costs of reducing class sizes to 18 students in Years 1 to 3 in the United States, as President Clinton has proposed, will require hiring an additional 100,000 teachers at a cost of $US5–6 billion per year. Per student costs are about $US500 for each year the students are in smaller classes. To reduce again from 18 to 15 students would cost a further $US5–6 billion per year. There is also the costs of classroom space, changes to buildings, teacher training, and so on. This investment could, instead, be used to raise teachers' salaries by $20,000 per year.

In summary, the research on class size indicates that very small gains are made to achievement as a consequence of reducing class sizes (even down to 15). Overall, the gains, where they exist, are positive, which has led to much advocacy for reducing class sizes. The costs of reducing class sizes to such an extent are extremely large, and it is likely that these resources could be more effectively used to achieve greater achievement gains, and higher quality teaching performances by implementing alternative innovations. There is also little evidence to claim that more and/or positive peer interactions accompany the benefits in achievement until class sizes are reduced to around 15 students.

4.3 Composite Classes

Composite classes are classes that include students of more than one year level who are taught in the same classroom by the same teacher. As mentioned in the introduction to this chapter, composite classes are the norm in New Zealand primary schools. In part, the high incidence of composite classes is due to the large number of small schools in New Zealand, especially rural schools. But even larger schools also use this form of organisation as a way of handling uneven numbers of students at different year levels. Schools also use composite classes because they are believed to have certain pedagogical advantages over single-level classes. According to Trussel-Cullen (1994), "they allow for more flexible grouping and learning styles, they encourage children to help each other and work together cooperatively and collaboratively, and they present more of a 'family' or 'community' atmosphere" (p. 30). So, whereas some schools use composite classes out of necessity, many choose to use composite classes as a matter of policy.

Other terms used to refer to this type of class are 'multigrade', 'multi-age', 'combination', 'split-grade', 'vertically grouped', 'mixed-aged', 'family group', and 'non-graded', although these terms are not all synonymous. 'Multigrade', 'combination', or 'split-grade' are terms frequently used in the United States and Canada to refer to classes in which students of more than year level are taught together for administrative or economic reasons (e.g., to even out class numbers). 'Multi-age' and 'family grouped' are terms used in the United States and Canada to refer to classes where students of different ages are put together for perceived pedagogical benefits. 'Mixed-age' or 'vertically grouped' are terms used in the United Kingdom, although they seem to refer to both multigrade and multi-age classes (in New Zealand, the term 'vertically grouped' is used to refer to classes comprising students of at least three Year levels). 'Non-graded' is a term applied in the United States to a department or school-wide programme where students are flexibly arranged according to their academic performance, rather than age, and proceed through the levels at their own rates (see Goodlad & Anderson, 1987). Non-graded programmes include the Joplin Plan, already discussed (for a meta-analysis of the effects of non-graded programmes on student achievement, see Gutierrez & Slavin, 1992). Multigrade and multi-age come closest to what New Zealand (and Australian) educators refer to as 'composite' classes, so this section focuses on the effects of multigrade and multi-age classes and their equivalents.

Veenman (1995) conducted a best-evidence synthesis of the cognitive and affective outcomes of multigrade and multi-age classes in primary schools across a variety of English-speaking and non-English-speaking countries (New Zealand was not represented). His review is exemplary in the comprehensiveness of the literature reviewed and the careful attention to criteria for study inclusion (these included initial comparability of samples and at least two experimental and two control teachers involved in the studies). In reviewing 34 studies comparing multigrade and single-grade classes and eight studies comparing multi-age and single-age classes, Veenman found no differences in achievement (median effect sizes of .00 and –.03 for multigrade and multi-age classes, respectively). However, in 13 studies of multigrade classes and eight studies of multi-age classes, he found small effects on students' attitudes towards school, self-concept, and personal adjustment favouring these classes (median effect sizes = .10 and .15, respectively). There was some variation, albeit inconsistent, in outcomes, depending on study quality, but little variation in outcomes by Year level or academic area (reading, maths, language). As a consequence, Veenman concluded that "parents, teachers, and administrators need not worry about the academic progress or social-emotional adjustment of students in multigrade or multi-age classes. These classes are simply no worse, and simply no better, than single-grade or single-age classes" (p. 367). Veenman also noted that, although few studies provided information on the instructional practices used in the classes, those that did suggested that teachers rarely capitalised on the multigrade or multi-age arrangement to promote learning from peers (e.g., by using cooperative learning or reciprocal teaching). Nor did teachers group students within the classes across grade or age lines in order to tailor instruction to more homogeneous groups (e.g., by using some form of Joplin Plan).

Mason and Burns (1996) criticised Veenman's conclusion, arguing that his null finding for multigrade classes is an artefact of selection bias favouring these classes, combined with lower quality instruction, which counteracts the benefits of selection. They argued that multigrade classes generally have better students and perhaps better teachers and that these selection factors mask a small negative effect resulting from the increased demands on teachers due to the greater diversity of students. Support for this claim comes from several interview studies with principals and teachers in multigrade classes in California (Burns & Mason, 1995; Mason & Burns, 1995, 1998; Mason & Doepner, 1998). Mason and Burns (1996) hypothesised that, when student and teacher selection factors are controlled, comparative studies of achievement in multigrade and single-grade classes should show a small negative effect in the order of –.10 of a standard deviation. Some support for this claim comes from their review of comparative studies involving small, rural schools where purposeful assignment of teachers and students is not possible (Mason & Burns, 1997a, 1997b). In their response to Veenman, Mason and Burns (1996) also argued that, because of the additional time demands placed on teachers in multigrade classes, teachers might neglect non-core subjects such as science and social studies, which would lead to negative effects on achievement in these areas.

In a reply to this criticism, Veenman (1996) reported results of a reanalysis, using meta-analytic procedures, of 51 multigrade and 12 multi-age studies, incorporating all studies from his best-evidence synthesis plus seven additional studies and one sub-study. Overall, the results again showed no significant differences in either cognitive or affective outcomes between multigrade/multi-age classes and single-grade/single-age classes. The effect sizes were essentially zero for cognitive outcomes and slightly positive, but still close to zero, for affective outcomes. The reanalysis showed some country-to-country variation in effects of multigrade classes on cognitive outcomes; there were small positive effects for studies conducted in the United States (mean effect size = .05) and Canada (.08) and a small negative effect for studies conducted in Europe (–.05). It also showed a small positive effect of multigrade classes for students in Years 1 to 3 (mean effect size = .06), a near-zero effect for Years 4 to 5 (.01), and a small negative effect for Years 6 to 7 (–.08). There was some support for the possibility that student achievement may suffer in subjects such as science (mean effects size = –.19) and maths (–.25). But there was no support for the notion that there might be small negative effects in schools where selection factors would not be operative (in rural schools the mean effect size was .10).

Nevertheless, there seems to be some agreement between Veenman and Mason and Burns that teachers rarely capitalise on multigrade or multi-age arrangements to promote learning from peers. Instead, teachers tend to teach distinctly different curricula, maintain grade levels, and deliver separate lessons to each grade-level group (Mason & Burns, 1997a, 1997b; Veeman, 1995). In an excellent study of mathematics achievement, Mason and Good (1996) compared the curriculum, instruction, and organisational formats used by primary school teachers in six multigrade classes with those used by teachers in 18 single-grade classes – six who used whole-class teaching and 12 who used two within-class ability groups. They coded 153 lessons taught by these teachers according to classroom type, the manner in which the teachers organised students for mathematics, and the nature of teacher-directed and independent-group activities. Teachers of multigrade classes organised their students into two groups for almost all lessons. Moreover, in independent group activities, students in the multigrade classes were less productive than were those in the single-grade classes, even compared to those that used a similar two-group structure. Students in the multigrade classes seldom worked cooperatively to solve problems and seldom helped others who were in need of assistance. Mason and Good noted that, whereas multigrade classes might provide opportunities for teachers to use more innovative, developmental approaches, these data provide little support for this notion. There was no evidence of increased opportunities for social growth, peer tutoring, and independent learning for students in the multigrade classes.

Unfortunately, the 'jury' is still out on whether multigrade or multi-age classes have effects, be they positive or negative, on students' learning outcomes. Whichever way the argument turns, the evidence suggests that any effects are very small – no more than a tenth of a standard deviation. The evidence on affective outcomes is clearer in that there are small positive effects, more so for multi-age than for multigrade classes. The reasons for this finding are not well documented, and student selection factors could still be at work (see Mason & Burns, 1997a, 1997b). There are claims that there is a greater incidence of giving help and sharing, greater social responsibility, and greater sensitivity to others in multi-age classrooms (Chase & Doan, 1994), but, as indicated by the Mason and Good (1996) study, there is evidence that contradicts these claims. Clearly, the success (or lack thereof) of multigrade and multi-age classes is moderated by the nature of teachers' instruction.

4.4 Single-sex Classes

As well as the debate about single-sex schools (see Chapter 5), there is from time to time a resurgence of interest in establishing single-sex classes within coeducational schools. Much of the interest comes from writers exhorting the advantages that would accrue from these classes for girls (Milligan & Thomson, 1992; Parker, 1985; Willis & Kenway, 1986), citing the differential nature of teacher interactions, intimidation of girls by boys, marking and assessment bias, and the content and presentation of subjects. There have also been numerous accounts written in popular magazines that have boosted these claims (see Willis & Kenway, 1986).

A common argument in favour of single-sex classes is that teachers give more time to boys in coeducational classes, hence disadvantaging girls. This is confounded with the nature of activities in the classes, with all-girl classes more likely to be confined to lower-order intellectual activities, whereas all-boy classes are more likely to engage in higher-order intellectual activities (Kenway & Willis, 1986; Wilce, 1984a, 1984b). These findings suggest an effect more related to teacher perceptions. Gillibrand, Robinson, Brawn, and Osborn (1999) investigated the reasons as to why 47 of a class of 58 girls choose to enter a single-sex class for physics (taught entirely by males). The major reasons were expectations of better results, avoidance of disruption from boys, wish to be with friends, and desire to experience the novelty. The major reasons for girls choosing mixed-sex classes, on the other hand, were that all-girl classes were demeaning and that boys helped them with their work. Kruse, in an extensive series of studies in Denmark (Kruse, 1987a, 1987b, 1989a, 1989b, 1990, 1991, 1992, 1994a, 1994b, 1994c, 1994d, 1995, 1996), reported that solidarity can be strengthened within girls classes, while the competitive element which often worked in favour of the 'attractive' boys was diminished. Parker and Rennie (1997) reported on the Single-Sex Education Pilot Project, initiated in 1992 in Western Australia, which was aimed at increasing girls' performance in mathematics and physical sciences via single-sex classes. There was no random assignment to classes. The teachers perceived that single-sex classes benefited those girls who were experiencing a great deal of harassment from boys in mixed-sex classes, although there was least benefit for the higher-achieving girls and boys. There was less willingness to teach boys-only classes, although this unwillingness dissipated during the project. The overall conclusion was that the effects were more dependent on the teacher and teacher expectations than whether the class was mixed- or single-sex.

There are very few defensible evaluations of the differences in learning outcomes for these studies. One of the major difficulties in addressing the impact on student learning from comparing students in single-sex compared with coeducational classes has been the problem associated with the non-equivalent group comparisons. Single-sex classes tend to be more selective both in students and teachers, and it is not clear whether these selection factors rather than the gender of the student could account for any differences. For example, at the school level, Steedman (1983, 1984) examined the achievement differences for a large sample of 16-year-old students in single-sex schools. After correcting for achievement levels before the students entered the school, he concluded that "very little in these examination results is explained by whether schools are mixed or single-sex once allowance has been made for differences at intake" (p. 98).

Rowe (1988) has conducted the most powerful study of single- compared with mixed-sex classes, as he was able randomise the students and teachers to the classes. He conducted his study at one high school in Victoria, Australia, where 189 students in Year 7 and 209 in Year 8 were randomly assigned to six single-sex or two mixed-sex classes for mathematics. The conclusion of the first report of his study (Rowe, 1988; Rowe, Nix, & Tepper, 1986) was that "both boys and girls allocated to single-sex classes gained higher levels of mathematics achievement, and demonstrated more positive shifts in attitudes, than those of their counterparts in mixed-sex classes" (p. 13). Rowe (1988) reported a two-year follow-up on these students. There were complications in that timetabling constraints necessitated that 111 students changed out of their randomly assigned classes; and missing data (inevitable in studies over two years) meant that only 261 of the original 398 students had complete scores. For those students who remained in the randomly assigned classes for the duration, and for those who shifted from single- to mixed-sex (or vice versa), he reported that gender differences were not statistically significant for either mathematics achievement or mathematics confidence across the years of the study. He did find, however, that students from single-sex classes were more likely to choose higher-level maths courses, and this he attributed to higher maths confidence scores in the last year of the study. "These results suggest that class-type effects on student confidence alone were sufficient to predict Year 10 course choice" (p. 195). Hence, Rowe concluded that, "by any criterion, the overall findings from the present study indicate that the institution of single-sex mathematics classes at the school studied has been a success" (p. 196).

Marsh and Rowe (1996) reanalysed these data using only students who had no missing data over the three testing occasions (pre-test, post-test, and two years later). Across all measures, there were no instances of gains for girls in girls-only classes being significantly more positive than gains for girls in mixed-sex classes (similarly, for boys classes). Hence, there was "no support for the advantages of single-sex mathematics classes for either boys or girls" (p. 153). Further reanalyses of the choice of higher-level maths classes, using more sophisticated methods of analysis, also demonstrated that there were no effects from the choice of class on subsequent maths choices. Marsh and Rowe did find that brighter students benefited more from being in mixed-sex classes and, where there were differences, these seemed to favour mixed-sex rather than single-sex classes. Similarly, Signorella, Frieze, and Hershey (1996) completed a 10-year longitudinal study of single- and mixed-sex classes within one private school, and concluded that there was "no consistent tendency for students in single-sex classrooms to display less gender stereotyping… [and there was] no consistent advantage to girls in single-sex as compared to mixed-sex classes" (p. 606).

Overall, there is very little compelling evidence of a compositional effect related to whether a class is single- or mixed-sex. It needs to be noted that most studies have been conducted on high school students and there is minimal research on these classes at the primary level; although we see little reason to suspect that there would be meaningful differences at this level. There are more powerful effects due to the quality of teaching and teacher expectations than to whether a class is all one sex or mixed.

4.5 Implicating Peer Effects on Learning

Given the major finding of this chapter, that very few meaningful compositional effects accrue directly from different classroom configurations, it is appropriate to suggest that such configurations have few implications for peer effects on learning. This is particularly underlined when it is also noted that changing configurations is unlikely to be accompanied by changes in instructional methods. The rhetoric that typically accompanies advocacy for changing classroom configurations includes enhanced peer interactions and more collaborative peer learning opportunities, but this rhetoric is not realised in practice.

It is likely that any compositional effects from changing classroom configurations have more influence on attitudes than on achievement. Most powerfully, streaming, for example, reinforces low expectations for both teachers and students and has differential effects on enjoyment of school – lower for low streams and higher for high streams. Streaming does tend to polarise student attitudes into pro- and anti-school camps (Gamoran & Berends, 1987): "Whereas high-track students tend to accept the school's demands as the normative definition of behaviour, low-track students resist the school's rules and may even attempt to subvert them" (p. 427). This polarisation is often echoed in the students' perceptions about their peers as well. The 'haves' become less likely to want to 'associate' with their less well-endowed peers and the 'have nots' are less likely to want to 'associate' with those with privilege to opportunities seemingly denied to them – by ability, by social stratification, and by institutional practices. Both high- and low-stream students, like their teachers, view the top streams as offering a better education, more opportunity, and more prestige. The earlier, particularly British, research found that a student's friends are found in the same stream (Ball, 1981; Cottle, 1974; Hargreaves, 1967; Rosenbaum, 1976; Scribner & Cole, 1981), with students in the upper streams tending to be the most popular in the school. Further, high-stream students supported each other in their class work, whereas low-stream students made derogatory remarks towards those who made academic efforts and competed against others in the same stream. As Ogbu (1974) found, a low grade (say 'C') was considered to be good enough for the lower-stream students; few students aspired to anything higher, which tended to keep students from advancing up the streams.

In our model of peer influences on learning, there is no reason to suggest that either the ambient environment or the frequency of tutorially configured experiences necessarily changes as a consequence of classroom configuration. There is much advocacy that changing classroom configurations *should* change the opportunities to implement the various tutorially configured experiences, and that the homogeneity effects of achievement within these classes would allow for an enhanced ambient environment and increased frequency of the various mechanisms. These opportunities seem to be rarely realised. The mechanisms of peer influences are no more likely to occur in higher- than lower- streamed classes, in smaller than larger classes, or in single-sex than mixed-sex classes. Other more powerful influences dominate classroom configurations.

4.6 Relevance to New Zealand

With the exception of research on composite classrooms, there is remarkably little research on effects of different classroom configurations in New Zealand – despite the prevalence of many of these configurations. We have noted above the widespread prevalence of streaming and composite classes, and the variability in class sizes across and within schools. There appears, however, to be few single-sex classes within co-educational schools in New Zealand.

With respect to streaming, Kealy (1984), in a qualitative study of Auckland Grammar School, reported that the effects of streaming for those in the lower-ability classes did not produce a supportive learning environment. Instead, these lower-streamed classes enforced a non-academic stereotype, which prevented any overt display of diligence and thus increased disadvantage (see also Elley, 1984).

There has been more extensive research on the effects of composite classes. Research on multigrade and multi-age classes may not be directly translatable to composite classes in New Zealand. Teachers in composite classes in New Zealand do not teach along 'grade' lines (as in multigrade classes in the United States), nor do schools use composite classes solely for pedagogical purposes (as in multi-age classes).

Recent research in New Zealand suggests that students in composite classes perform somewhat less well than do students in single-level classes. In an HLM analysis of New Zealand's performance in the 1990 IEA study of reading literacy, Wilkinson (1998) found that nine-year-old students in composite classes performed significantly less well in reading comprehension – by 2.06 percentage points – than their counterparts in single-level classes. This difference translates into an effect size of –.11. This finding contradicts that of Chamberlain's (1993) analysis of the IEA data, showing no significant difference in achievement between composite and single-level classes. The effect observed by Wilkinson was net of other factors such as the socio-economic standing of the school (there is a tendency for more composite classes to be found in schools serving more multicultural and lower socio-economic communities). Similarly, Wylie et al. (1999), in their report of data from the 523 children participating in the Competent Children project at age eight, noted that students from composite classes had somewhat lower scores in several literacy- and numeracy-related competencies than students in single-level classes. Not all of these differences could be explained by school socio-economic status.

Wilkinson and Hamilton (1998) examined the reasons why composite classes might lead to lower performance of students, at least in reading. Following on from Mason and Burns (1996), they proposed that the greater diversity of students in composite classes might stretch teachers' capacities to cater for students' instructional needs. Hence, they proposed, students may be given less direct and intensive support, or less precisely tailored support, for literacy learning. Using data from 484 students in 18 classes in nine primary schools, they compared the range of abilities in comparable composite and single-level classes and how teachers accommodated the diversity of instructional needs in these classes. Contrary to expectations, results showed that the range of abilities in composite classes was not necessarily greater than that in single-level classes. Moreover, regardless of whether a class was composite or single-level, or the diversity of abilities within the class, some teachers seemed better able than others to structure their teaching to accommodate students' needs. Wilkinson and Hamilton interpreted these findings as evidence against causal explanations for achievement differences between composite and single-level classes. They suggested that findings from previous studies of small negative effects of composite classes may simply be an artefact of the mix of year levels from which students in the samples were drawn and the operation of selection factors that result in biased estimates of student performance. In the absence of any other explanation, they concluded that composite classes in New Zealand did not appear to contribute to lower achievement, at least as far as reading was concerned.

Although there are many articles advocating smaller classes for New Zealand schools, there is a paucity of evidence about the consequences of this innovation. Podmore (1998; Podmore & Craig, 1989) in a New Zealand study concluded that, when class size is reduced, other more macro-level changes need to occur, such as changing value systems and promotion practices, before learning outcomes gains can be accrued (see also McDonald, 1988).

4.7 Conclusion

The research in this chapter shows that there is a small advantage for many of the classroom configurations. The best estimates of the effects of streaming on achievement outcomes is .05, of class size (from 30 to 15 students) .13, of composite classes (perhaps) –.10, and of single-sex classes .05. Over all the various effects, at best the effects average about .10, and this estimate only increases when teachers change their instruction to more fully adapt to the students in their classes. This change does not mean changing the pace of instruction or lowering the expectations of what the students can accomplish, but a dramatic change in the nature of the activities, a renewed vigour towards implementing appropriately challenging tasks, and implementing the many positive mechanisms that lead to enhanced student learning.

As we noted above, the trade-off is not between closing the gap between low-ability and high-ability students versus raising overall student test scores via implementing various class-level configurations. Rather, it is between policy makers attending to classroom organisation practices versus improving what happens once the classroom door is closed. Whether a school streams or not, reducing class sizes, and/or implements composite or single-sex classes appears less consequential than whether it attends to the nature and quality of instruction in the class, whatever the within-class variability of achievement. It is almost certain that there are conditions of learning, such as those identified in Chapter 2, that are far more powerful. Good teaching is more powerful and can be independent of the class configuration or homogeneity of the students within the class.

The most important implication of the research in this chapter is that the major cost of attending to class-level configurations relates to the false belief that something educationally sound has been accomplished. By changing the homogeneity of student achievement at the class level, by implementing single-sex classes, by reducing class size, or by creating composite classes, there is little change in the probability that peers will be implicated in the learning process, or that students will have enhanced learning outcomes. Nothing, or at best, very little, will change. This is the cost of attending to class-level configurations.

The attention needs to be directed at more careful curriculum specification, higher quality teaching, and higher expectations that students can meet appropriate challenges – and these occur once the classroom door is closed and not by reorganising which students are behind those doors.

4.8 Recommendations for Further Research

- There is need for research on streaming in the New Zealand context, with particular attention to the effects on students in lower streams. Given the above research evidence, it is important also to research why teachers and schools persist with streaming in the face of the consistent message that it makes little difference and may lead to teachers and principals (and parents) believing that its presence makes a positive difference.
- This research should attend to both the achievement and equity outcomes, but concentrate more on the nature and quality of teaching and peer experiences in these various classroom configurations. The current research points to the effects of lower teacher expectations, attention primarily to pacing, and the nature of the teacher-student interactions as more powerful. It is not clear how peers interact to improve one another's learning, if at all, as a function of these class configurations, other than teachers and students changing their expectations.
- Before implementing a class-level change, it may be most valuable to research the effects of differing professional development programmes to optimise the perceived benefits supposedly associated with the implementation. Given that most positive effects noted in this chapter relate to the importance of changing teaching practices rather than student composition, the sustained benefits of such professional development and careful and systematic evaluation may point to the manner in which to capitalise on these class-level effects.
- Given the advantages, albeit minor, of reducing class size to around 15 students, particularly in the first few years of schooling, careful cost-benefit analyses of this implementation relative to other innovations need to be investigated.
- As there is a large number of composite classes in New Zealand and their overall effects might possibly be negative (albeit small, about –.10), more research needs to be undertaken on how teachers could change their instruction (probably to teach across age, tailor instruction to provide more innovative and developmental approaches, and increase opportunities for social growth, peer tutoring, and independent learning for all students in the composite classes). Such research seems imperative given the large number of such classes in New Zealand schools.
- There is also need for research in terms of teaching classes with concentrations of students by ethnicity and culture.

CHAPTER 5

SCHOOL COMPOSITION

This chapter discusses the literature on the impact of between-school differences in student intake, as indexed by measures of school composition, on student achievement. This literature indicates little consensus over the nature and size of school compositional effects despite over 30 years of research, so we give considerable attention to discussion of likely problems within the literature. Most of the research in this area has been concerned with the composition of student intakes in terms of ability or socio-economic status, but the gender and ethnic makeup of schools has also received some attention.

New Zealand communities and their schools vary widely in their characteristics. In terms of ability composition, the best estimate is that approximately 16% of the total variance in student achievement in New Zealand is due to differences between schools (Postlethwaite & Ross, 1992). By comparison, 18% to 19% of total variance in student achievement is due to differences between schools in the United States (Lee & Bryk, 1989a; Postlethwaite & Ross, 1992). There are also large differences in the socio-economic composition of New Zealand schools as indicated by the Ministry of Education's decile system (Ministry of Education, 1996) as well as by studies using the Elley-Irving (1985) index (e.g. Lauder & Hughes, 1990; Thrupp, 1997a). In a study of one New Zealand city, Lauder et al. (1994) found levels of residential socio-economic status and ethnic segregation that were comparable to New York, though not as great as Chicago or Los Angeles. There are also considerable differences between schools in ethnic composition, not only between kura kaupapa Maori and most state schools, but also among state schools generally (Ministry of Education, 1996).

In 1998, 85% (2,341) of all schools in New Zealand were state schools, four percent (117) were private schools, and 11% (312) were integrated schools ('private' schools that receive state funding for teacher salaries and operational cost but no provision for capital works or building maintenance, as the property is owned by the school, not the state) (Ministry of Education, 1999c). Integrated schools retain their right to select students, charge fees to cover their capital costs, and have a slightly higher decile rating than state schools (Wylie, 1998). In 1995, 26% of boys and 25% of girls attended single-sex secondary schools (Nash & Harker, 1998). Single-sex secondary schools in New Zealand are more selective than coeducational schools, with Maori students less likely to attend than non-Maori and students from low socio-economic backgrounds less likely to attend than students from high socio-economic backgrounds (Nash & Harker, 1998). In 1996, 0.7% of school-age students were home schooled (Ministry of Education, 1996). The number of students receiving home schooling has increased considerably in recent years, with a 250% increase in student numbers since 1989 (Kerslake, Murrow, & Lange, 1998). In 1998, the average roll size for New Zealand primary schools was 185 students; the average roll size for intermediate schools was 413; and the average roll size for secondary schools was 711 (Ministry of Education, 1999c).

In the first section of this chapter, we examine the statistical analyses of school compositional effects from studies over the last decade. In the second section, we turn our attention to the literature on school compositional effects at the 'extremes' of different student compositions; here we compare the achievement of students in single-sex versus coeducational schools and in public versus private schools, and explore the effects of home schooling. In the third section, we examine the implication of school size as a factor that might moderate school compositional effects. The fourth section looks at how research on school composition might implicate peer effects on learning. The last sections conclude and point to areas where further research is required.

5.1 School Composition Broadly Defined

The effect of school composition on student achievement has been the cause of debate for some time because it appears that large-scale empirical studies have produced very little consensus on the issue. Broadly, there are three types of explanation for this. One approach, which we call the 'conceptual and statistical issues' explanation, would invoke a wide range of conceptual and statistical problems associated with the literature to explain the lack of consensus. A second explanation, which we call the 'insufficiently detailed modelling' explanation, would argue that the HLM/MLM studies that have investigated school composition are incapable of capturing them. Finally, a 'political' explanation is put forward by Thrupp (1995, 1999b), who, in reviewing the history of research in this area, has argued that the debate has been dogged by ideological and political considerations, since the outcomes of key policy decisions (e.g., school choice, effectiveness, and accountability) depend on the findings.

Although the 'insufficiently detailed modelling' and 'political' explanations are important, and we shall have more to say about them in the conclusion to this chapter, in this review we work within the more immediate terms of the 'conceptual and statistical issues' explanation. We are concerned, in what follows, to evaluate empirical studies on school compositional effects over the past 10 years, seeking to provide a meta-analysis that goes beyond the crude summation of tests of significance, as, for example, undertaken by Hanushek (1986a). In order to provide such an evaluation, however, we need to negotiate a series of complex conceptual and statistical issues. We adopt the following strategy. In the next section (5.1.1), we examine some of problems involved in measuring school compositional effects and, as a result of this discussion, develop an 'ideal' model of the variables and design needed to establish the nature and significance, if any, of school compositional effects. In subsequent sections (5.1.2 – 5.1.3), we evaluate recent empirical studies in the light of this 'ideal' model. In section 5.1.4, we make a judgement as to the impact of school compositional effects based on the analysis contained in the earlier sections. In section 5.1.5, we comment on findings from New Zealand research on school effects.

5.1.1 Conceptual and statistical issues involved in measuring school compositional effects

The theorisation of school compositional effects has evolved historically. First brought to prominence by Coleman and colleagues (Coleman et al., 1966), school compositional effects were then conceived as a hypothesis about the impact of peers on motivation, aspirations, and attitudes towards education. Almost two decades later, Barr and Dreeben (1983) published their influential study of primary schools, which showed that the characteristics of a student group significantly influenced the teacher's work. In a review of research of high school organisation and its effects on students and teachers, Bryk, Lee, and Smith (1990) reported that, "the overall distribution of student characteristics shapes the curricular offerings of a school and the policies which map students into courses" (p. 147). What these extensions of the original school compositional effects hypothesis suggest is that the effects can have a far wider effect on school organisation and performance than originally conceived by Coleman et al. (1966) and by much of the literature on this subject since that date (for a review of this literature, see Thrupp, 1999b). School compositional effects could involve not only 'peer group processes' (in the strict sense of student reference group processes), but also 'instructional' and 'school organisational and management processes' as well.

The most comprehensive statement of school compositional effects is that of Thrupp (Thrupp, 1999a). Although, in an ideal world, we would be seeking to identify studies that incorporated all three dimensions (peer group processes, instructional, and school organisational and management processes) into their research design, these studies are few and far between. Thrupp's (1999b) qualitative study is one of the few that does so (for a description of this study, see section 5.4 of this chapter). The lack of studies incorporating all three dimensions may be, in part, because many researchers have not been aware of the existence of this model and, in part, because in quantitative terms it is clearly difficult to find proxy measures for the processes involved in curriculum organisation and school policy. When it has been attempted – most imaginatively by Chubb and Moe (1990) – it has run into difficulties (see section 5.1.3). With these caveats in mind, we turn to a closer examination of the issues involved.

A key problem is that of complexity. The work of Coleman et al. (1966) provides a good example of the problems involved. They raised the possibility that, although school compositional effects could have a significant impact on school performance, it might also be the case that ethnic minority or working-class students might suffer from low self-esteem in schools with predominantly middle-class students. A similar argument could be mounted in relation to student sub-cultures; that is, in socially well-mixed schools, the effects of school composition would be cancelled out by student sub-cultures in which those of high prior achievement excelled, while 'those of lower prior achievement generated a culture of resistance and school failure. Given these hypotheses, the HLM/MLM studies that might most clearly demonstrate school compositional effects would be at the extremes of intake (i.e. those schools with solidly high or low socio-economic status intakes).

A number of further issues follow from this. First, what do HLM/MLM studies tell us about the performance of students from different social class or ethnic backgrounds in reasonably well-mixed schools? If there is a whole-school effect, is this because school policy has managed to 'integrate' working class/ethnic minority students into a whole-school culture of academic performance? If so, then we would expect school compositional effects to show up. Conversely, where these students are not integrated, we would not expect to see school compositional effects. Second, the argument suggests that only studies with a wide coverage of schools (i.e., high socio-economic status, well-mixed, and low socio-economic status schools in the same sample) would be able to inform the debate over school compositional effects. Third, longitudinal studies would also throw light on the impact of school composition. This is because, although some variation in school performance could be expected from year to year and also from department to department in secondary schools (Sammons, Thomas, & Mortimore, 1997), the school compositional effects hypothesis would suggest that school performance would not fluctuate dramatically, unless the intake also fluctuated.

There are, then, some potentially sound reasons why there might be disagreement over the impact of school composition, because it is suggested that school compositional effects would only exist under certain circumstances and could only be tested by studies which conform to the specifications outlined above. Ideally, the sample that would best establish the strength of school compositional effects would be one that examined the performance of schools at either end of the socio-economic spectrum (in New Zealand terms, decile 1 and 10 schools). A similar point is made by Sammons et al. (Sammons et al., 1997) when they say:

> *An alternative explanation which may influence the statistical significance of school context variables may be found in the range and extent of advantaged versus disadvantaged intakes in the total sample of schools. In other words, contextual effects may be more likely in LEAs [Local Education Authorities] or regions where a policy of selection is employed by a minority of schools (p40).*

The reason for this is that, where a small number of schools operate under a policy of selection by high prior achievement, high socio-economic status students are likely to be creamed off by these schools, thereby polarising intakes. An effect similar to this can be seen in the work of Harker and Nash (1996) in which a school compositional effect associated with students' prior achievement was identified in two selective, high socio-economic schools.

Complexity is also created by the need to measure school compositional effects in particular types of schools. The issue of whether school compositional effects can be identified in primary schools is important because it addresses the hypothesis articulated by Lauder et al. (1999) that school outcome effects at the secondary level can be strongly predicted by prior achievement. In this hypothesis, primary schools will have been influential, especially if there are school compositional effects at work in them. It is also worth noting that many of the large samples from the United States include analyses of the Catholic sector. Here the difficulty is that there may be a self-selection effect, because particular groups of parents may positively opt for their children to go to Catholic schools. Nevertheless, studies such as those of Bryk, Lee, and Holland (1993) are instructive in that they compare the performance of public (state) and Catholic schools.

Another set of problems revolves around measurement issues. One concerns the type of school compositional effects to be measured. This is the question as to whether measures of ability and/or prior achievement are used or whether some measure of social class is preferable. One uncontroversial finding from the school effectiveness literature is that measures of 'ability' or prior achievement are the best predictors of future outcomes. However, prior achievement composition is rarely entered as a variable in these studies, as we shall see. Rather, some measure of social class or socio-economic composition is entered. However, while we would not expect a perfect correlation between measures of prior achievement and socio-economic status, the two are clearly related. Therefore, ideally, studies seeking to measure the impact of school compositional effects should be constructing a range of compositional variables rather than simply one, as is often the case. In addition, they should indicate the degree of correlation between variables such as socio-economic status and prior achievement (to indicate he degree to which one variable is serving as a proxy for the other).

A further measurement problem concerns the ability of these variables to successfully measure school mix. The traditional measure of school composition used in the literature is that of the mean socio-economic status and occasionally the mean ability or prior achievement. However, there is a view that parental education levels and especially mothers' education would provide a more powerful predictor of a compositional effect. Of course, there is a debate about the extent to which parents' education reflects their social class position, and, again, studies should ideally assess the degree to which they are correlated. However, at best, these indicators are likely to be weak proxies for compositional processes and effects, if we take the three-dimensional model of school composition into account.

Finally, there is an issue of interpretation. In most quantitative studies, there are, as we suggested above, difficulties in disentangling causation in relation to school processes. In discussing this point, Lauder and Hughes (1990) pointed out that:

> *...the data on school mix can, in theory, be interpreted according to at least three models. In the first model, the effects associated with school mix reflect unmeasured intake characteristics.... In the second model, the variation in attainment associated with school mix is a proxy for variations in unmeasured school processes.... The correlation between school mix and successful school processes could come about because (i) successful processes cause school mix, (ii) school mix causes successful school processes, or (iii) some third variable causes both.... The third model identifies school mix as a genuine causal variable (p50).*

Given these difficulties in interpretation, we need to need to be cautious about what it is that school compositional effects represent even when they are identified. For example, in relation to the first model, Thomas and Mortimore (1996) have suggested that social class composition may in fact be a proxy for prior achievement variables. This is a point to which we shall return. In the case of the second model, the evidence presented by Lauder et al. (1999) in relation to flows of students in secondary markets in New Zealand suggests that parental choice is exercised – largely, although not exclusively, in terms of social class – rather than successful schools attracting higher levels of composition in any straightforward sense. In other words, there is evidence that can be brought to bear in assessing these different models. This is not just a matter of arguing for a favoured explanation on political or ideological grounds.

In light of the above discussion, we can now delineate a model, within the conceptual and statistical issues explanation, against which we can assess the studies that have examined school composition over the past decade. First, the sample should include schools from both ends of the socio-economic spectrum. School compositional effects are unlikely to appear in reasonably well-mixed schools. Ideally, the samples should draw only from both ends of the spectrum (rather than from the middle). Second, a full set of entry-level variables, including prior achievement variables, need to be included to establish whether compositional variables are acting as proxies for other variables, especially prior achievement. Third, there should be measures that can capture the possible correlations between the three dimensions of the school composition model (peer group processes, instructional, and school organisational and management processes). Fourth, a combination of compositional variables (e.g., prior achievement mix or socio-economic composition) should be constructed and the relationships between them identified. These would include mean socio-economic status measures and measures of parental education. Fifth, where possible, a mix of school types would be included in the sample (e.g., Catholic and non-Catholic, primary and secondary schools). Sixth, where possible the study should be longitudinal. Seventh, additional data should be collected to narrow down the possible interpretations of compositional effects given the three models of interpretation outlined above. Finally, in addition to the above, we assume that studies should conduct their analyses according to multi-level modelling techniques.

With these criteria in mind, we are now in a position, within the conceptual and statistical issues framework, to evaluate the studies of the past 10 years. We have divided the literature into two clusters, British studies and United States studies, because, as we shall see, there are important differences between the two. They typically use different compositional measures and vary in the degree to which they can illuminate the relationships between school composition, curriculum, and policy. On the other hand, while the disciplinary background of the studies will often be mentioned, there is little to be gained from dividing up the literature by disciplines, as nearly all the studies are 'production function' studies of one kind or another and nearly all use HLM or ML. In reviewing the British and United States literature, we begin in each case with a general overview before examining key studies in more depth. The examination of the key studies does not pretend to be exhaustive; we do claim, however, to bring out the key issues and dimensions to the problems. It also reviews the most high profile and/or substantial contributions to the area.

5.1.2 British research on school composition

Over the past decade, a number of British school effectiveness studies have found little or no evidence of school compositional effects. For instance, Mortimore, Sammons, Stoll, Lewis, and Ecob et al. (1988) found only a weak effect of socio-economic status, and Thomas and Mortimore (1994; Thomas & Mortimore, 1996) claim that, when rich and wide-ranging data on prior achievement are available at the pupil level, the effects of school composition disappear. Gray, Jesson and Sime (1990), in a study of 11 Local Education Authorities (LEAs), Bondi (1991), in an analysis of reading attainment in Scottish primary schools, and Strand (Strand, 1998), in an analysis of primary school performance across an English LEA, found little evidence of school compositional effects either. On the other hand, some school effectiveness research interest in school composition has been reflected in a concern with ability composition, or what Mortimore (1995) calls an 'academic balance'. Smith and Tomlinson's (1989) study of multi-racial British comprehensive schools did not test for socio-economic status effects but found a weak effect of mean ability on the general progress of students, as did Maughan and Rutter (1987), whose analysis does not discuss effects of socio-economic status at all.

In British sociology of education, the view that school composition is a key theoretical construct was given particular prominence in the late 1980s by the findings of McPherson and Willms (1987), who examined the effects of creating a system of comprehensive schooling in Scotland between 1970 and 1984. They found that 'comprehensivisation' – that is, creating the comprehensive schooling system – significantly reduced social class inequalities of attainment and improved average levels of attainment when measured against the inequitable pattern established in the preceding six decades. Following earlier work by Willms (1986), McPherson and Willms (1987) attributed the decline of social class inequality in attainment to a school compositional effect. They found that comprehensivisation resulted in the abolition of selection at the age of 12, the closure of many short-course schools, and the redefinition of school catchments, which, they argued:

> *...led to a reduction in between-school segregation in many communities. This reduction, allied to the rise in the socio-economic status level of the school population, distributed the*

benefits of a favourable school context more widely, though it must be added that these benefits are not well understood (p23).

More recently, Willms has suggested that "the composition of a school's intake can have a substantial effect on pupils' outcomes over and above the effects associated with pupils' individual ability and social class" (Willms, 1992, p. 410) (see also Rumberger & Willms, 1992; Willms & Raudenbusch, 1989). There is also support for school compositional effects among a number of other British researchers (Gibson & Asthana, 1998; Heath & Blakey, 1992; Patterson, 1991).

We now turn to a more detailed investigation of the British literature, focusing on the work of Smith and Tomlinson (1989), Gray, Jesson, and Syme (1990), Thomas and Mortimore (1996), Sammons, Thomas, and Mortimore (1997), Robertson and Symons (1996), and Strand (1998).

Smith and Tomlinson's (1989) studied 18 schools with a multi-ethnic intake selected from various areas within England. The mix hypothesis was tested in relation to reading, maths, and ethnic composition. Here, the results were not significant, although the prior achievement scores were "in the expected direction" (p. 278); that is, the correlations were close to significance. Measures of prior achievement were factored in. The authors' conclusion is that their findings "lend some limited support to [the proposition] that all children tend to do better at schools that have a high proportion of high-attaining children" (p279). Other compositional variables relating to social class and indeed unemployment (some schools had high level of unemployed parents) were not tested.

Gray, Jesson, and Syme (1990) examined the performance of schools within 11 LEAs, covering 14,000 pupils in 290 schools. However, prior achievement data were available for only three LEAs and none of these had information on social class. In the LEAs where they had social class information, they found that, once school composition on this variable was entered, the total variance in outcomes explained by between-school factors dropped by a quarter (i.e., from 16% to 12% of the total variance). However, in the three LEAs where measures of prior achievement could be entered at the individual level, "the evidence for strong compositional or contextual effects was weak" (p. 149). There are several points to make about this study. In the three LEAs in which there was a measure of prior achievement, the social class school compositional variable was not significant, suggesting that measures of prior achievement will 'cancel out' socio-economic compositional effects. A point to note is that the report provides no indication as to how the social class variable is constructed. However, as we understand it, there was also no prior achievement compositional variable constructed. The data for this study were not rich in either individual or school-level variables, although, in terms of sample size, it appears to fulfil the criteria for encompassing the full range of schools.

Thomas and Mortimore (1996) studied 11,881 secondary students in 87 schools in Lancashire. However, 28% of pupils had to be excluded because of incomplete prior attainment data. This group were more likely to be eligible for free school meals (FSM) and have lower achievement in examinations, suggesting that the estimates for this study would be "biased upwards" (p. 8). The major proportion of the reduction in school-level variance was due to the inclusion of prior achievement measures. In this, it is consistent with Gray et al. above. Several contextual variables were included in this study, although there was no direct measure of socio-economic status. Of these proxy measures, the most interesting were those relating to the qualifications of parents, which included the percentage of parents with higher qualifications and the percentage of parents who were unskilled. The proxy measures Thomas and Mortimore used for school composition were not significant when prior achievement data were entered. The authors call for further research on the issue on the influence of compositional factors, being aware that other studies had found these to be significant.

Sammons, Thomas, and Mortimore's (1997) study is one of the few with longitudinal data, using data over three years from 69 London schools. Proxy measures of social class, such as free school meals (FSM), were used, and data on ethnicity were also used, but the full range of possible compositional factors were not explored, since there was no prior achievement composition entered into the calculations. The authors found that, over three consecutive years, "school context is less significant than individual student factors, except for English [results], where the influence of socio-economic context seems to be particularly strong" (p. 42). One reason for this may be that there is a relationship between those on FSM and second language English speakers. These will largely be clustered into a few schools where such an effect would not be surprising. This would be consistent with our hypothesis that it is at the two ends of the spectrum that we could expect composition results to be significant.

Robertson and Symons (1996) examined data from the *National Child Development Study* (1958), including 18,359 students. The study looks at the influence of social composition and streaming in primary schools. The data analysed are drawn from various sweeps of this cohort and include measures of prior achievement at age seven and outcome measures (for reading and maths) at 11. The data show that students with high prior achievement did better in streams than did those in an unstreamed school. However, those of lower measured achievement did worse than predicted. They did better in mixed-ability situations in which schools were well mixed. The interesting point of this study is that it establishes a social class compositional effect and then compares students of varying ability in streamed and unstreamed situations.

Strand (1998) examined 1,698 students in 57 primary schools using an abstract reasoning measure of prior achievement. However, it was administered in the term prior to students taking the outcome measure, the *Key Stage 2* tests. In terms of assessing 'value added', this is clearly problematic. FSM and ethnicity data were included in the study, and it included the abstract reasoning measure as a compositional variable. In terms of composition, only an age variable was significant. The study is interesting for highlighting the issue of age, something that is rarely done in school effectiveness studies. However, its utility in relation to illuminating the school compositional effects debate is low, because of the timing of the value-added measure.

Overall, the evidence provided by these studies is clearly mixed, yet none really approximates our ideal model either. The most significant point is that the British studies do not have a complete measure of socio-economic status; rather, a proxy measure, such as percentage of free school meals, is used. There is both a measurement and theoretical problem associated with such proxies. In terms of measurement, the percentage of FSM tells us nothing about the remaining socio-economic composition of the school, so it is only where there is a very high proportion of free school meal places in a school that we can be reasonably confident that we have an appropriate measure of socio-economic school mix. The theoretical problem concerns what in the United States is called the 'tipping hypothesis'. Namely, that a school's performance declines once a certain percentage of students living in poverty, or from lower socio-economic groups, or from ethnic minority groups enters a school. But this hypothesis can only be fully tested if we have complete data on socio-economic background. In other words, the tipping hypothesis may be dependent not only on the percentage of disadvantaged students in a school, but also on the nature of the remainder of students in a school.

The most reliable of the English studies would be that of Sammons, Thomas, and Mortimore (1997) because of its longitudinal nature, and the most interesting in relation to school composition and streaming would be that of Robertson and Symons (1996) because it addresses the links between school composition and school organisation. This, for example, is not an issue that any of the other studies addresses. Yet, according to the composition hypothesis articulated earlier, it is important to identify studies that throw some light on this relationship. To make further headway, we need to consider the studies carried out in the United States.

5.1.3 United States research on school composition

In the United States, school effectiveness 'sensitivity to context' research has raised the possibility of stronger school compositional effects by highlighting the limitations of a comprehensive 'recipe' approach to effectiveness in schools with different intake characteristics. In an often-cited study, Hallinger and Murphy (1986) found that, for the most part, schools of different socio-economic status have quite different effectiveness correlates. "High and low socio-economic status effective schools [are] characterised by different patterns of curricular breadth, time allocation, goal emphasis, instructional leadership, opportunities for student reward, expectations for student achievement, and home-school relations" (Hallinger & Murphy, 1986, p. 347). Teddlie et al. (1989) and Teddlie and Stringfield (1993) reached similar conclusions. One reason for the differences between United States and British school effectiveness researchers on this issue may be the tendency for United States researchers to study more wide-ranging samples that highlight contextual issues, whereas British researchers tend to study more homogeneous, socially disadvantaged schools (Reynolds, 1992; Reynolds, Sammons, Stoll, Barber, & Hillman, 1996).

In relation to the sociology of education in the United States, Pong (1998) has argued:

> *In 1966, Coleman et al. found that schoolmates affect students' academic achievement. Subsequent explorations of the impact of school composition led to near unanimity that schools' average socio-economic status matters for all its students. That is, sociologists generally accept that schools with greater percentages of students from high socio-economic status family backgrounds provide more effective learning environments and lead to higher academic achievement for all (p23).*

This may overstate the support for school compositional effects among United States sociologists, however. Certainly, there is much United States research evidence of school compositional effects associated with socio-economic status and ability (see Bryk et al., 1990; Bryk et al., 1993; Caldas & Bankston, 1997; Gamoran, 1992; Lee & Bryk, 1989a; Roscigno & Ainsworth-Darnell, 1999 for helpful reviews) and some evidence of effects of race (Caldas & Bankston, 1997; Mahard & Crain, 1983; Roscigno, 1999; Roscigno & Ainsworth-Darnell, 1999). At the same time, Gamoran (1987a, 1987b) and Bryk and Driscoll (1988) both find counterbalancing compositional effects of socio-economic status and ability, while only two studies suggest little or no evidence of school compositional effects of any kind (Jencks & Mayer, 1990; Lee & Smith, 1995).

United States economists and political scientists working in education have also often been interested in peer effects because of the way students are regarded as co-producers; that is, part of the resource base of schools as well as the consumers. Although this literature tends to give school composition much more emphasis than, say, the school effectiveness literature does, the findings are still extremely varied. Evans et al. (1992) found little evidence of school compositional effects. However, they are careful to point out that their study does not argue that peer group effects are inconsequential but that the central issue is methodological problems in the identification of school compositional effects. Link and Mulligan (1991) also report only weak effects of school composition. Some involved in the debates over public versus private schooling, vouchers, and school choice argue that effects of school composition are relatively insignificant (e.g., Chubb & Moe, 1990; Evans & Schwab, 1995), but others argue that school compositional effects are important (e.g., Epple & Romano, 1998; Levin, 1998; Zimmer & Toma, in press).

We turn once again to a more detailed examination of the literature. The studies reviewed are Chubb and Moe (1990), Bryk, Lee, and Holland (1993), Ho and Willms (1996), Pong (1998), and Zimmer and Toma (in press). With the exception of Zimmer and Toma (in press), these studies go further than the British studies in illuminating the relationships between school composition, curriculum, and policy.

Chubb and Moe (1990) draw upon two samples, one drawn from the HSB 1980 sophomore cohort and the other drawn from the 1984 HSB Administrator and Teacher Survey (ATS). The data from these surveys were not gathered over the same period of time, and this raises a question about the relationship between measures such as school composition and school organisation, since the measures of school composition were taken from the HSB sample and measures of school organisation from the ATS survey. Moreover, because the authors seek to demonstrate that school organisation is more significant than school composition, they do not examine the possible relationships between them. Rather, they make a set of causal assumptions, for which they have been taken to task, about the primacy of school organisation in determining school outcomes. Nevertheless – and this may be considered an irony – in their full sample of private and state schools, they find school composition in terms of socio-economic status to be significant. Arguably, this is a more robust finding than those relating to school organisation.

Bryk, Lee, and Holland's (1993) study is probably the most interesting study to emerge from the United States in relation to the school composition hypothesis. It is a book-length study based on HSB data set, which compares the performance of Catholic and non-Catholic state schools. The authors found a strong effect of school social class on average achievement but the effects are different in the two sectors. They also found that when these compositional or contextual effects are taken into account, Catholic schools perform better. However, when maths is the dependent variable and the number of maths courses taken is entered, the compositional effects are largely washed out. The inference drawn by the authors from this is that the "differential pattern of school social compositional effects for public and Catholic schools... appears to work largely through basic features of the organisation of schools (overall size and differentiation in students' courses of study and their normative environment)" (p. 268). Bryk et al. (1993) findings that "minimizing disciplinary problems is a necessary condition for the routine pursuit of academic work" (p. 270) concur with those of Thrupp (1999b). The more orderly nature of Catholic schools, and the curriculum and normative environment they are able to construct, are clearly related to the parent-school relationship in Catholic schools which Bryk et al. describe as "distinctive". That is, they are a self-selecting subset, which enables compositional effects to be mediated in Catholic schools, thus explaining their superior performance. For the purposes of this review, this is probably the most robust and interesting statistical study to have been published, because it provides strong inferential evidence of the links between school composition and school organisation. We should also note, however, the methodological point that composition needs to be entered into the analysis ahead of organisational factors such as curriculum options, otherwise the compositional factors are obscured.

Ho and Willms's (1996) study utilises the NELS sample of 24,599 eighth-grade (Year 9) students comprising a representative sample of schools in the United States. It seeks to examine the relationship between four dimensions of parent involvement in schools and their children's performance. Although measures of socio-economic status, family structure, ethnicity, and any learning and behavioural problems constitute the family and student background variables, there is no measure of prior achievement. A mean socio-economic status, school compositional variable was constructed. The authors find that the effects of parental involvement are mediated by the school compositional effect. This effect was extremely powerful: "Children scored considerably higher in both mathematics and reading if they attended a high SES school, irrespective of their own family backgrounds" (p. 138). The study is of real interest because of the relationship it establishes between school composition and the nature of parental involvement. However, the study would have been more powerful if it had been able to include a prior achievement variable.

The study of Pong (1998) also utilises the NELS sample. This study sought to examine the role of family structure, especially single-parent families, in relation to exam success. However, the study also takes into account measures of socio-economic status. In addition, a measure of social capital was included as measured by parental social relations, as was school type. The author was therefore able to discriminate between family and socio-economic effects. She also included students' Year 9 IRT (Item Response Theory) 'estimated number right' scores as a measure of prior achievement, and the dependent variables were Year 11 mathematics and reading test scores and IRT scores. The most interesting aspect of this study is the author's attempt to test the tipping hypothesis discussed above, with the proportions of students from single-parent families in each school taken into account. The author finds that, "although the school-level variables (minority status, socio-economic status, and parental relations) substantially reduced the negative influence of a high concentration of students from single-parent families, an effect still remained" (p. 38). The author also observes that schools with high concentrations of single-parent family students are so low in economic and interpersonal relations that such schools may fail to attract good teachers. This then suggests that there is a spillover effect into school recruitment of teachers and other personnel.

The work of Zimmer and Toma (in press) is also of considerable interest because it is, to our knowledge, the first to examine the question of a school compositional effect in an international comparative context in which school type is also taken into account. The countries in the study are the United States, France, Belgium, Ontario (Canada), and New Zealand. The data sets for these countries, taken from the 1981 IEA study of mathematics achievement, included measures of prior achievement, socio-economic background of students, the attributes of the students' teachers, and scores of other students in the classroom. The authors find that the higher the mean test score of class mates, the higher the achievement level of students. Like Summers and Wolfe (1977), they also find that students of lower prior achievement are likely to do better the higher the achievement of their classmates. Where high- and low-ability students are mixed in a classroom, most of the gains are captured by low-ability students, the estimated 'peer effect' being .52. In addition, mixing ability types leads to higher student achievement. The authors conclude that, "on average across countries and school types, increasing the class mean on the beginning of the year's scores, increases a representative student's end-of-year test score by approximately .25" (Zimmer & Toma, in press, p. 15). Stated alternatively, moving a student from a class with a mean test score of 15 to a class with a mean score of 25 will increase the student's achievement level by 2.5 points. When fathers' and mothers' occupations and education levels of classmates were entered along with measures of prior achievement, fathers' education and mothers' education were both significant. Again, it was found that improving the mix of fathers' education benefited lower-ability students more than those of higher ability. Overall, the authors conclude that "peer effects appear to characterise school production across institutional settings...; the institutional arrangement for schooling that maximises the benefits from peers should be considered further" (pp. 22–23). Although this is clearly interesting, it does raise a question about the level of the analysis. An inference is made from classroom-level data to the school level. However, in some respects, this seems an appropriate inference because various dimensions of the compositional effect relative to various ability levels are explored. This suggests, other things being equal, that there are gains to be made by optimising the mix of students in relation to parental educational levels and prior achievement levels. What the study cannot do is throw light on the mechanisms at work to produce this effect, especially at the school level.

Overall, the United States studies over the past decade show the existence of a school compositional effect, irrespective of whether they include prior achievement measures (and only one of the studies reviewed does not include a prior achievement measure). These studies explore a range of possible interaction effects between school composition and levels of ability and parental participation. However, only one study – that of Bryk, Lee, and Holland (1993) – seeks to directly link school composition to school organisation, and here it is by a process of inference.

5.1.4 Summary of selected British and United States studies

The balance of evidence from the literature reviewed in Britain and the United States demonstrates a school compositional effect when judged against the criteria set out earlier in this chapter. For presentational purposes, Table 5.1 summarises the analysis according to the variables by which this review has evaluated the studies. The best of the British studies reports a strong compositional effect in relation to English, and the most recent study of primary school performance also reports a compositional effect. From the United States, the evidence is stronger, although the familiar caveat in relation to American school-effectiveness studies should be noted – their dependent variables are rarely exams that seek to test curriculum knowledge. Only one study provides an estimate of the effect size of school composition – that of Zimmer and Toma (in press). And only one study seeks to explore the full school compositional effect hypothesis (Bryk, Lee, & Holland, 1993).

Table 5.1

Summary of British and United States Studies Reviewed

Study	Shows mix effect	Longitudinal	Prior achievement	Measure x3 dimensions	Covers range of social spectrum	School type
Smith & Tomlinson (1989)	no	no	yes	no	yes	no
Gray et al (1990)	no	no	yes	no	yes	NS
Thomas & Mortimore (1994)	no	no	yes	no	yes	yes
Sammons et al (1997)	yes	yes	yes	no	yes	yes
Robertson & Symons (1996)	yes	no	yes	no	yes	NS*
Strand (1998)	no	no	yes**	no	yes	NS
Chubb & Moe (1990)	yes	no	yes	no	yes	yes
Bryk, Lee & Holland (1993)	yes	no	yes	yes	yes	yes
Ho & Willms (1996)	yes	no	no	no	yes	no
Pong (1998)	yes	no	yes	no	yes	yes
Zimmer & Toma (in press)	yes	no	yes	no	yes	yes

*NS= Not stated.

** This study measure of prior achievement was shortly before the dependent variable was measured.

5.15 New Zealand research

The recent evidence from New Zealand is also contradictory. Three school-effectiveness studies have used HLM to investigate how much of the total variance in achievement may be apportioned between student-level variables and school-level, including compositional, variables (Harker & Nash, 1996; Lauder et al., 1999; Wilkinson, 1998). The inconsistencies in findings from these studies parallel those found in the overseas literature.

Analysis of compositional effects was only incidental to Wilkinson's (1998) study. He conducted an analysis of nine-year-old students' performance in the 1990 IEA survey of reading literacy. The analysis used data from a representative sample of 173 coeducational primary schools throughout New Zealand, and the major outcome variable was a composite measure of students' comprehension of narrative, expository, and document text. Differences between schools accounted for 16.98% of the total variance in students' comprehension – a figure almost identical to that of Postlethwaite and Ross (1992) – and class or school composition (mean socio-economic status of students, proportion of girls, and proportion of students for whom English was a second language) accounted for a significant proportion of the between-school variance over and above that due to the corresponding individual-level factors (individual students' socio-economic status, gender, and home language). Unfortunately, no estimates of the proportions of total or between-school variance explained by the class- or school-composition factors were provided. Moreover, the results almost certainly overestimate the compositional effects, because there was no measure of prior achievement in the student-level model. .

Harker and Nash (1996) used data from the Progress at School Project (Nash & Harker, 1998, 1997) to conduct a value-added analysis of effective schools. The sample was 5,393 students attending 37 secondary school throughout New Zealand, representing approximately a 10% sample of the annual population cohort, and the outcome variables were School Certificate marks in English, maths, and science. The schools were selected from a group of 50 schools that had expressed a willingness to participate in a project to investigate school effects on student performance. There were 25 coeducational, six single-sex boys, and six single-sex girls schools, including two private schools. The schools were equally distributed among large and middle-size cities and small towns, and they differed in the proportions of Maori and Pacific Islands students enrolled. Two 'elite' boys schools (one of which was private) were regarded as outliers, since they contributed over half the between-school variance in maths achievement, so they were excluded from the analysis. Harker and Nash found that differences between schools accounted for only five to nine percent of the total variance in students' achievement, and that school composition (mean initial ability, mean socio-economic status, and proportions of Maori, Pacific Islanders, and Asians) accounted for little between-school variance beyond that explained by the corresponding individual-level variables. Specifically, school composition accounted for only about one to two percent of additional between-school variance in maths and science, and adding compositional variables slightly **reduced** the between-school variance accounted for in English (an artifact of the high leverage exerted by particular schools in the small sample). Hence, school composition explained essentially zero percent of the total variance in student achievement. On the basis of these results, Nash and Harker (1997) conclude:

> *The model suggests that once the character of a school's intake has been taken into account no systematic differences can be detected in the performance of schools when School Certificate marks are used as the criterion. The hypothesis that the 'ability' or social class composition of a school has an independent effect on a school's performance is shown to be doubtful (p. 5).*

Lauder et al. (1999) analysed data from the Smithfield Project to assess the impact of various New Zealand Government reforms on a cohort of 3,300 students as they moved from Year 7 (towards the end of primary) to Year 11 (third year of secondary). In the main sample of this study, students from 23 schools were selected, including four single-sex boys, three single-sex girls, and 16 coeducational schools. Nineteen schools were state schools. Ten schools were predominantly Pakeha, three were predominantly Maori and Pacific Islands, and 10 were of mixed ethnicity. Eight schools were classified as upper socio-economic status, eight were middle, and seven were lower socio-economic status. The HLM analyses actually involved two sub-studies: one at Year 10 using students' scores on standardised tests of study skills, vocabulary, and verbal analogies (a self-concept scale was also used, but it did not prove informative in the analysis); and the other at Year 11 using students' School Certificate results in English, maths, and science. Lauder and Hughes found that, on average, differences between schools accounted for 16.15% of total variance in the outcome measures, and school-level variables accounted for approximately 45% of the between-school variance beyond that accounted for by individual-level variables (specifically, the school variables explained 55% of the between-school variance in standardised test scores and 34% of the between-school variance in School Certificate results). The school variables included a wide array of compositional variables (mean prior achievement scores, mean aptitude, mean reading comprehension, mean socio-economic status, mean scores on the self-concept scale, and proportions of students in various ethnic groups) and other school characteristics (e.g., size, stability of roll, and numbers of positive teacher comments). Taken together, the school-level variables were calculated to account for about eight percent of the total variance in student achievement. These results led Lauder et al. (1999) to conclude:

> *The mean SES of the students in the school, their mean prior achievement scores, and the like are related to performance over and above the relationships found at the individual level. Schools with larger proportions of students with high initial achievement, larger proportions of students with high socio-economic status, fewer ethnic minority students, stable rolls, and the like are at an advantage, and students will perform better in them than they will in schools with the opposite mix of students (p. 127).*

In an earlier publication, based on the same database, Lauder and Hughes (1990) demonstrated that it was primarily socio-economic mix and not school type (private, single-sex) that accounted for the most variability in school outcomes.

How can differences in findings from these two studies be reconciled? In our judgement, Harker and Nash's (1996) analysis underestimates the effects of school composition. If we assume that 16% of the total variance in student achievement is due to differences between schools in New Zealand, and this estimate seems right, then Harker and Nash's analysis grossly underestimates the between-school variance that is potentially explainable by school factors. This is probably attributable to a small, non-random sample of schools, possibly restricted in range of ability and/or socio-economic status, that may not be representative of the country as a whole. The underestimation is also probably attributable to Harker and Nash's use of School Certificate results. A considerable amount of self (i.e., student) and school selection goes into whether and when students sit School Certificate. This selection process may serve to partial out much of the between-school, as well as within-school, variance in achievement. Added to this, Harker and Nash adopted a technically thorough and very conservative approach to their analysis (indeed, their study is exemplary for the technical sophistication of the use of HLM methodology). They excluded two outlier schools because of the gender imbalance this introduced in the achievement distribution. And they chose to retain all student-level variables in their level-1 models, even those variables that made non-significant contributions to prediction of students' performance, in order to make fair comparisons between schools. They also used a limited set of compositional variables. The combined influence of these factors makes it unlikely that strong school compositional effects could be found.

Conversely, we judge that Lauder et al.'s (1999) analysis overestimates the effects of school composition. Curiously, despite an even smaller sample of schools, they appear to have estimated the appropriate amount of total variance that is due to differences between schools in New Zealand. This might be because they had a more representative sample of schools, representing a wider range of ability and/or socio-economic status, or it might be because they used more stringent criterion measures (or both). Lauder and Hughes used students' scores from several standardised tests, but they also estimated students' School Certificate results in a way that may have minimised the bias inherent in this measure. They used clever estimation routines to calculate School Certificate results for those students who had already sat the Sixth Form Certificate equivalent, or who had not completed School Certificate but had completed the corresponding unit standards. Nevertheless, we believe that their estimate of eight percent of total variance is an overestimate because it relates to other school-level variables, in addition to compositional variables, and because they excluded non-significant student-level variables from the level-1 models. Moreover, some effects cannot be interpreted as true compositional effects because the corresponding individual-level variables were not included in the models. This applies to the proportion of Pacific Islands students in the vocabulary model; proportion of Maori students and mean reading comprehension scores in the English model; and mean reading comprehension scores in the maths model. Lauder and Hughes also used a much wider array of compositional variables than did Harker and Nash.

Overall, the discrepancies in the New Zealand studies reflect those observed elsewhere for much the same reasons. In our judgement, the proportion of total variance in student achievement that is attributable to school composition in New Zealand lies within the range of zero to eight percent. Based on our interpretation of Lauder et al.'s (1999) analysis, we guess it is just under four percent. The inconsistencies in results are probably due to sampling error, different criterion measures, and different modelling of compositional effects. Larger samples of schools and better outcomes measures (not School Certificate results), comparable to those available in the large national databases used by United States researchers, are essential if New Zealand researchers are to derive robust estimates of the effects of school composition (cf. footnote 64 Wylie et al., 1999).

5.2 Particular Dimensions of School Composition

Other research has examined effects of school composition in schools at the 'extremes' of different student compositions. In this section, we review literature that compares the achievement of students in public versus private schools, single-sex versus coeducational schools, and home schooling versus public or private schools.

5.2.1 Public versus private schools

Too often, research has treated private schools as a uniform 'black box'. Private schools, however, vary considerably and are defined differently in different countries. At minimum, they are schools that are not government owned. Most often, they are more selective than public schools, take students from higher socio-economic backgrounds, and have more latitude over their running and accountability (Wylie, 1998) – but this is a norm, not a prescription or definition. It is widely assumed that private schools offer students a superior quality of education to that of state schools, and hence have generated much debate about access, particularly via vouchers (Wylie, 1998).

This issue of 'superior quality' is hotly debated. In a recent review, Witte (1996) observed, "there is no question [that], in terms of raw scores, private schools have better test results. The issue is whether these higher scores can be explained by controlling for prior achievement, student and family characteristics, school context and any selection effects" (p. 164). The findings of most research comparing public and private schools are questionable, according to Witte (1996), as they do not control for selection effects, particularly those relating to prior achievement.

Some of the problems inherent in studies using HSB data are discussed in a review of research in this area by Jencks (1985), who outlined the debate stemming from the study by Coleman, Hoffer, and Kilgore (1982). Based on analysis of the 1980 HSB data, Coleman and his colleagues concluded that Catholic school students learned considerably more than public school students during their last two years at secondary school. However, Alexander and Pallas (1983) and Willms (1983) reanalysed the data and noted that the benefits identified by Coleman and his colleagues were "considerably smaller, statistically insignificant, and a product of specification error" (Jencks, 1985, p. 128). Jencks noted there is agreement between the three studies on three main issues: (i) Year 9 scores in 1980 are higher in Catholic schools than in public schools; (ii) between Year 9 and Year 11, the raw scores increased more for Catholic school students than for public school students; and (iii) after controlling for Year 9 student characteristics, Catholic school students still showed a larger gain than public school students on reading, vocabulary, maths, and writing tests. However, all three studies produced differing estimates of sector effects. Jencks identifies three sources of variation between the studies – the choice of Year 9 control variables, the sample of students selected for analysis, and the tests emphasised in each study. According to Jencks, four conclusions can be drawn from these studies. First, Year 9 and Year 11 students learned slightly more in Catholic schools than in public schools. Second, the magnitudes of effects were uncertain; the point estimates averaged .03 or .04 standard deviation per year in the study, but the confidence interval for the population value was broad and the effect size varied by test. Third, the evidence for Catholic school benefits for disadvantaged students was persuasive but inconclusive. Fourth, the cumulative effect of attending a Catholic school for twelve years was only three times the effect of attending a Catholic school for the past two years. This may be because Catholic elementary schools are less distinct in character than Catholic high schools or because the benefits from elementary schools become weaker over time.

Chubb and Moe (1990) have also undertaken comparisons of the public and private school sector in the United States. Chubb and Moe have been instrumental in the increase in private school provision in the United States and have encouraged the use of vouchers as a means of increasing student achievement. In their research, they found that private school students had an advantage compared to those in public schools, and they attributed this to greater autonomy of the schools in the private sector. However, their findings have been widely questioned. For example, Glass (1997) states that Chubb and Moe failed to consider HSB data that show that teachers in the two sectors rate their autonomy no differently for many aspects of classroom-based decisions. In his own study, Glass found size and cost to be the main differences between the two sectors, with private schools spending twice as much per student than schools in the public sector. He concluded that national legislation and limited funding restrict the autonomy of teachers in both sectors. Glass further criticises Chubb and Moe's claim that school autonomy contributes to student achievement. Glass and Mathews

(1991) point out it is more likely that student achievement fosters autonomy rather than the other way around. Smith and Meier (1995) also question Chubb and Moe's use of student-level data to assess effectiveness at the school level. They used state-level and district-level data in their analysis and considered changes over time, and found little support for Chubb and Moe's conclusions. Bryk and Lee (1992) also identify several theoretical and methodological problems in Chubb and Moe's analysis. They note, for example, that the analytical method used when comparing scores of students in Year 9 and Year 11 resulted in higher estimated gains for initially higher-achieving students. They suggest that this served to increase the difference between effective and ineffective schools to the advantage of the private school sector.

In a recent review of research comparing public and private schools, Witte (1996) concluded that the majority of research findings indicate only small effects of private schools. Witte also noted exceptions to these findings. For example, analyses by Plank et al. (1993) using NELS data showed positive effects of attending private schools on both a Year 11 composite achievement test (beta weight = 1.13; se = .43) and a maths test (beta weight = 0.67, se = .22). They used a simple regression model, controlling for Year 9 scores, race, family background, student educational aspirations, and school location. Attendance at private school and public school of choice (those allowing selection of students) was coded by dummy variables. However, as noted by Witte, the significance levels of coefficients and standard deviations of dependent variables were not mentioned. As such, it is not possible to assess the magnitude of the effects.

Goldhaber (1996) estimated Year 9 NELS test scores for maths and reading, with separate regression models for public, Catholic, and private (non-Catholic) schools. Goldhaber's model included a comprehensive set of student-, classroom-, and school-level variables, and a dummy variable for a learning disability or being held back a year. His model also included variables for the number of courses a student had taken in the subject area, and the usual family variables. Unmeasured selection effects were also controlled for. Goldhaber calculated what a given student would have achieved in Year 11 had that student had the same school characteristics and attended a school in an alternative sector. Results showed that nearly all achievement differences disappeared. The analyses showed that the majority of differences between sectors were due to student differences and that, once they were controlled for, there were even negative sector effects for private schools. For example, in reading, the difference dropped from .51 standard deviations to –.07, making the reading achievement score higher for public school students. A public school advantage was also found for maths. Goldhaber's analysis also found that private school students in the lowest quartile gained the most, whereas students in the top quartile were disadvantaged. This finding is consistent with other evidence on this issue.

Research by Evans and Schwab (1993a) has produced very different results to Goldhaber (1996). Evans and Schwab used HSB data and focused on high school completion and subsequent enrolment in college. After adjusting for ethnicity, gender, age and religious attendance, family income and structure, and parents' education, they calculated that a representative student in a Catholic school retains a .12 advantage in the probability of graduating from secondary school and a .14 advantage in the probability of attending college. These results proved robust to the introduction of all other variables including third-form test results, school peer variables, and family education resources available. Consistent with a growing trend in research findings, this study also found effects were stronger for lower-achieving students (as reported by Goldhaber). These results are also consistent with earlier research that found dropout rates are very low in Catholic high schools – only about one quarter as high as in public schools (Coleman & Hoffer, 1987). Dropout rates for minorities in Catholic schools have been found to be considerably lower than for their counterparts in public schools (Bryk et al., 1993).

One explanation for positive effects of attending a private school is what has come to be known as communal school organisation. Bryk, Lee, and Holland (1993) identify three key components to a communally organised school. The first component is the presence of shared values about the purpose of school, what students learn, and how teachers and students should behave. The second component is shared activities, both academic and non-academic, that provide opportunities for face-to-face encounters among community members. The third component is the presence of social relations based on an ethos of caring. Central to this is the expansive role for teachers and collegiality between staff. The joint existence of these three components, Bryk et al. claim, creates a school life that powerfully affects its members.

Comparing Catholic and public high schools, Bryk, Lee, and Holland (1993) investigated the consequences of communal school organisation on several outcome measures of teachers' commitment and students' engagement. The sample consisted of 340 schools (8,650 teachers and 9,633 students) from the HSB data set. A comparison of Catholic and public schools on 23 indicators of communal school organisation revealed large Catholic school effects (over two standard deviations for the composite measure). Results showed that all effects favoured Catholic schools, and that effects on some indicators were large (one standard deviation or more). For example, for a percentage of students in extracurricular activities an effect size of 1.46 favouring Catholic schools was found, and for participation by teachers in faculty social events an effect size of 1.63 was found, favouring Catholic schools. This study provides solid empirical evidence that Catholic schools are more strongly characterised by a communal organisation than are public schools.

Further, Bryk, Lee, and Holland (1993) used a prediction equation to assess the likely improvement in teacher commitment and student engagement that would result for the average public school if that school were organised in the same way as the average Catholic school. The effects were found to be large for three teacher outcomes – efficacy, enjoyment, and staff morale (over .9 of a standard deviation) – and three student outcomes – missing class, classroom disorder, and interests in academics (approximately .5 of a standard deviation). They also computed a measure of the relative improvement that would occur in a typical school if it had the same level of communal organisation as the average private Catholic school. According to Bryk, Lee, and Holland, a more communal organisation would raise the typical public school from the 50th to the 80th percentile on teacher outcomes (efficacy, enjoyment, and staff moral). Student interest in academics would rise to the 66th percentile. On measures of student disengagement, the average public school's percentile would drop from the 50th to the 30th percentile for missing class and classroom disorder, and to the 37th percentile for dropout rate.

Bryk, Lee, and Holland (1993) also suggest that an inspirational philosophy has a positive, causal relationship to student achievement. They identify Christian 'personalism' and 'subsidiarity' as the two main ideas forming the basis of this ideology, which sets these schools apart from the public school sector. Crucial to Christian 'personalism' is the extended role of the teacher that encourages staff to care about the kind of people students become as well as the skills and knowledge they acquire. 'Subsidiarity' is based on considerations about work efficiency and specialisation and the mediation of this by a concern for human dignity. At the core of subsidiarity is a belief that the full potential of individuals is actualised in the social cohesion that can form around small-group associations. The suggestion that there is a causal relationship between inspirational ideology and student outcomes is, however, debatable, as there are no variables in the HSB data set to capture this concept directly.

Witte (1996) claims that the communal concept, so central to Bryk, Lee, and Holland's studies, has been poorly represented in their models that estimate achievement, and appears to be confounded with measures of academic climate, as such direct links between school characteristics and student achievement are weak. Despite these limitations, Witte is supportive of Bryk, Lee, and Holland's claims, as he considers the theory has certain logic and he concurs that Catholic schools appear to be superior in their ability to improve the performance of low-achieving students.

Other explanations for positive effects of Catholic schooling include a common core curriculum (Bryk et al., 1993) and higher academic focus (Witte, 1996). Evidence from studies using the HSB data indicate that private school students are more likely to be in an academic stream, to take more academically oriented courses, and to have higher academic aspirations. The evidence from studies using both HSB and NELS data suggest that these factors lead to greater achievement gains (Goldhaber, 1996; Plank, 1993). Bryk et al. suggest that a confined academic curriculum is a major contributor to the more equitable social distribution of achievement evident in Catholic high schools. Catholic high schools offer a core curriculum for all students, regardless of their personal background or future educational plans. Some students may begin the curriculum at a more advanced level and proceed in more depth, but the same basic academic goals apply to everyone. Although some streaming occurs, it is limited in scope and number of levels. This theory is supported by recent empirical work that links differences in student academic experiences to stratification in academic outcomes (Garet & Delany, 1988; Lee & Bryk, 1988; Lee & Smith, 1993).

In summary, student achievement is higher in private than public schools, but this effect can be explained by differences in school intake rather than effects of school type. However, there appear to be some positive effects for low-achieving students attending private schools. Furthermore, the research by Evans and Schwab (1993a) shows that Catholic school enrolment appears to have a strong effect on the probability that a student will complete high school and go to university. According to Witte (1996), Evans and Schwab's study is potentially important as "it indicates the most significant and robust positive effects to date of Catholic high schools" (p. 167). Bryk, Lee, and Holland suggest a common core curriculum, higher academic focus, a greater sense community, and an inspirational ideology are possible explanations for positive effects.

5.2.2 Single-sex versus coeducational schools

Traditional arguments for single-sex schools include a superior quality of education, higher standards of discipline, and the differing educational, social, moral, sporting, and cultural needs of boys and girls (Cocklin & Battersby, 1987). It is also claimed that coeducational schools are oriented towards a youth sub-culture where physical attractiveness and heterosexual popularity is seen to be more important by students than academic achievement (Coleman, 1961). Changing views and definitions of gender roles have led to new arguments favouring single-sex over coeducational schools, particularly for girls. It has been suggested that single-sex schools will provide more opportunity for girls to take leadership roles and will expose girls to strong female role models. It is also claimed that girls will be better off in single-sex schools because they will be free from harassment and teasing by boys. There is a body of research that documents the harassment of girls by boys in coeducational schools (Eder, 1995; Nash & Harker, 1998; Thrupp, 1999b). By contrast, those opposing sex segregation in education have claimed that coeducation enhances students' social adjustment and provides students with a more natural preparation for life (Cocklin & Battersby, 1987; Lee & Bryk, 1996).

There is renewed interest in comparing single-sex and coeducational education, particularly in the United States, because of a recent focus on secondary education and questions about school organisation (Lee & Bryk, 1986). Interest has also increased because of differing performance by boys and girls in some curriculum areas and the recent introduction of innovations such as single-sex classes within coeducational schools (see Marsh & Rowe, 1996; Rowe, 1988). Early research comparing single-sex and coeducational schools tended to focus on student attitudes about the social and psychological environments of their schools rather than on the impact on their academic performance. These studies generally found coeducational schools to be friendlier and more relaxed than single-sex schools, and single-sex schools to have a greater emphasis on control and discipline, especially for girls. In a review of this literature, Lee and Bryk (1986) caution that the results of earlier studies need to be considered in the context of a time when single-sex schools were increasingly viewed as having an oppressive influence on students' social development.

Historically, the most prominent research on achievement differences between these two types of schools is Dale's (1974) research in England. Dale found higher levels of achievement of boys in coeducational grammar schools than in single-sex grammar schools and no difference in achievement of girls between the two types of schools. On the basis of the results, Dale claimed that coeducational schools provided a superior environment for students' social and affective development and that the social and affective benefits of coeducation did not come at the cost of poorer academic performance.

Riordan (1985) examined achievement differences of white students between single-sex Catholic schools, coeducational Catholic schools, and coeducational public schools, using 1972 NELS data. Comparisons were made of student performance in each type of school on a range of variables including vocabulary, reading, and SAT (Scholastic Aptitude Test) verbal and maths scores. Riordan concluded that, on average, Catholic single-sex schools were nearly twice as effective with respect to achievement as Catholic coeducational schools, especially for girls, even after controlling for students' gender, race, socio-economic status, and geographic region. In explaining these findings, Riordan supported the view expressed by Coleman, Hoffer, and Kilgore (1981) that single-sex schools are more effective in achievement than coeducational Catholic schools, because there is a reduced adolescent subculture, fewer non-academic distractions, and fewer control and discipline problems. Furthermore, Riordan claimed that students benefit from being exposed to positive same-sex role models. However, these findings have been criticised for the limited statistical control of students' background characteristics (see Lee & Bryk, 1986).

Using data from the HSB survey, Lee and Bryk (1986) compared the effects of single-sex and coeducational secondary schooling on academic and non-academic outcomes. Their study is important as it is based on a large representative sample of students and provides the opportunity to study effects of school type on growth during the critical last two years of high school. They examined achievement differences for reading, mathematics, science, and writing for the two school types, using a random sample of students in Catholic high schools. The found that coeducational students did not surpass single-sex students in any achievement area at either Year 11 or Year 13. By contrast, there were statistically significant advantages from attending single-sex schools over coeducational schools, especially for girls. For boys, the benefits of single-sex schooling were larger at Year 11 (mean effect size = .17) than at Year 13 (mean effect size = .11), with no significant gains between Year 11 and Year 13. For girls, the benefits of single-sex schooling increased in size from Year 11 to Year 13 (mean effect size = .00 and = .11, respectively) and the gains in reading and science achievement were statistically significant (mean effect size = .14 and .20, respectively). Lee and Bryk considered the differences in school structure and resources between the two types of school as a possible explanation for their findings. Girls schools in the study had fewer students per teacher than coeducational schools, which may have resulted in a more intimate environment. Furthermore, single-sex schools for both genders offered a narrower range of courses, and the similarity of students' experiences in course taking may have contributed to the positive effects. Although narrow, courses in single-sex schools were more academic in nature and this may have produced a stronger academic focus.

Marsh (1989) discounted Lee and Bryk's findings because of a number of limitations in their methodology and study design. For example, the use of a one-tailed statistical test did not readily allow for significant effects to favour coeducational schools. As such, it was not possible to claim that there were no significant differences in favour of coeducational schools. Similarly, there were no controls in their study for pre-existing differences in academic achievement and other areas assessed as outcomes (the only outcome variable controlled for was educational aspirations). Marsh pointed to the extensive discussion by researchers in the field (for a review, see Jencks, 1985, April) where it is agreed that the only effective way of estimating effects of school type is through using both background variables and (in this case) 1980 scores to correct students' 1982 achievement test scores. Furthermore, Lee and Bryk claimed there was an interaction between school type and gender, but there is a lack of consideration of gender differences in their outcome variables.

Marsh (1989) conducted a further analysis of single-sex/coeducational differences for Catholic high schools using the HSB data. His study differed from that of Lee and Bryk (1986) in that he incorporated post-secondary outcomes based on the 1984 testing of the age cohort, and the inclusion of additional background and outcome measures. Marsh examined whether there were any effects of school type on 1982 and 1984 outcomes over and above differences due to background variables and 1980 outcomes. He also examined whether gender differences found in 1982 and 1984 outcomes related to effects of school type or to background variables and 1980 outcomes. This, claimed Marsh, provided the most legitimate basis for inferring effects of school type. His analyses revealed small but statistically significant differences between single-sex and coeducational students (e.g., beta weights of .06) and these tended to favour students from single-sex schools. However, the effects cannot be attributed solely to school type, as they may represent pre-existing differences – when Marsh introduced appropriate controls, the differences in 1982 and 1984 outcomes did not consistently favour students from single-sex or coeducational schools. Furthermore, for 14 of the 40 outcomes, effects of school type varied by students' gender.

More recently, LePore and Warren (1997) compared single-sex and coeducational Catholic schools using NELS 1988 data, with a sample of approximately 25,000 randomly selected students. The 1988 survey data were followed up in 1990, 1992, and 1994. The NELS 1988 data offer distinct advantages to data used by previous researchers, because the sample of Catholic single-sex and coeducational schools was large enough to enable sophisticated analyses comparing these schools. Furthermore, students were observed before they entered secondary school, which enables control for student differences prior to secondary school. LePore and Warren compared students' academic performance and social psychological outcomes between the two school types. They found that boys attending single-sex Catholic secondary schools scored higher on achievement tests than boys attending coeducational secondary schools. However, boys in both sectors increased their scores between Year 10 and Year 13 by approximately the same amount, indicating that boys in single-sex schools did not learn more. Girls in the two types of school had almost equal scores, and there were no significant differences in gain scores between single-sex and coeducational students. This indicates that students in single-sex schools did not learn any more than did students from the coeducational schools. No differences were found on social psychological factors such as locus of control and self-esteem (factors expected to be higher for girls in single-sex Catholic schools). Regression analyses showed that links between attending a single-sex school and students' achievement or psychological test scores were weak or non-existent. After controlling for difference in students' Year 9 achievement, there were no positive effects of attending a single-sex school.

In summary, research comparing achievement differences between single-sex and coeducational schools has produced mixed results. Several earlier studies based on the HSB data set claim effects for school type favouring single-sex schools, particularly for girls. However, these studies are plagued by lack of control for student background characteristics, since students were not assessed prior to secondary school, and single-sex schools are distinguished from coeducational schools by their tendency to enrol higher socio-economic status and higher-ability students. The more recent, more sophisticated studies, in which student background characteristics are less likely to undermine findings, show no consistent effects of school type favouring single-sex schools.

5.2.3 Home schooling

Another form of educational choice is home schooling. Home schooling refers to the provision of a student's education in the home by parents. The difference in the levels and type of social interaction in relation to learning between public and home-schooled students makes it interesting to compare the achievement of these two groups. To think that home-schooled students are completely isolated from the influence of peers is, however, a misconception.

Those who choose to home school have been accused of failing to provide children with the skills they need to cope in society (Arai, 1999). Implicit in these accusations is the view that schools provide this knowledge and parents are not able to. The knowledge and skills children are seen to be missing out on tend to relate to issues of socialisation and citizenship and fall outside the formal curriculum areas. It is argued that, in school, children learn how to work with others, handle conflict, and make sacrifices for others. It is claimed that home-schooled children are not sufficiently exposed to other people or the diversity of other cultures and will not be prepared for the realities of the competitive labour market. Parents who have decided to home school argue that the coping skills taught by school are simply consequences of their organisational structures. They also point out that socialisation within schools is limiting, as schools group children according to age so they do not gain the opportunity to learn from those younger and older than themselves. Furthermore, the level of conformity required by schools nullifies exposure to cultural diversity.

Research evidence suggests that when home-schooled children reach adulthood they are not only able to cope with life in wider society, but they cope very well (Meighan, 1995; Ray, 1994; Webb, 1989a). Contrary to arguments about home-schooled children's isolation and lack of exposure to others, research shows that home-schooled children are very involved in social activities outside the home (Knowles, 1998; Mayberry, Knowles, Ray, & Marlow, 1995; Ray, 1994; Thomas, 1998). Results of a survey by Ray (Ray, 1997) of 5,402 home-schooled students in the United States showed that these children engaged in an average of 5.2 activities outside of the home per week, with 92% of those surveyed involved in two or more activities.

The most substantial research to date comparing the achievement of home-schooled students and students in public or private school is a United States study by Rudner (1998) who surveyed and tested 20,760 home-schooled students and their families. Rudner found the achievement of home-schooled students was exceptional. Within each Year level and skill area, the median scores for home-schooled students fell between the 70th and 80th percentiles of students nationwide and between the 60th and 70th percentile of Catholic and private school students. For younger students, this represented a one-year advantage. By the time students were in Year 9, they were four years ahead of students in public or private schools. Home-schooled students were also found to perform well on the 1998 ACT (American College Test) college entrance examination, with average scores .38 standard deviations above the national ACT average, placing home-schooled students in the 65th percentile of all ACT test takers (Rudner, 1998). The findings by Rudner (1998) are supported by other United States studies (cf. Calvery, 1992; Ray, 1997). While Rudner's research shows home-schooled students outperform their counterparts in public and private schools, the home-schooling population is very select, with income and educational levels well above the national averages (Rudner, 1998). Furthermore, the act of home schooling sets this population apart in terms of their strong commitment to education and children.

In summary, research finds the achievement of home-schooled students to be exceptional. It must be noted, however, that there are substantial limitations to studies in this area in that they do not control for background differences between populations. As such, we are unable to draw strong conclusions with regard to comparative effects on learning.

5.2.4 Relevance to New Zealand

As mentioned previously, Lauder and Hughes (1990) conducted a comparison between public and private schools on students' achievement and occupational destinations on leaving school, using data from 20 schools in Christchurch. However, the results had more to do with the socio-economic composition of the schools than they did with the type of school. Unadjusted results showed an achievement advantage for students attending private schools. After controlling for intake characteristics such as gender, social class, and measured ability, the results still indicated a private school advantage. However, after controlling for socio-economic composition of the schools, the differences diminished, although taking socio-economic composition into account made little difference to the finding for destinations of school leavers for the different types of schools. Lauder and Hughes suggested that class-based student sub-cultures may be instrumental in explaining effects of school composition. They also suggest that, in a school with a balanced intake, the school ethos and mixing with students with the appropriate or valued cultural beliefs can serve to lift the performance of low socio-economic status students.

A small private school advantage (2.5 percentage points) was found in an analysis of the 1981 IEA study of mathematics achievement for 13-year-old children in New Zealand (Toma, 1996). However, this advantage was reduced when controls (based on parental occupation and education) were introduced for student socio-economic status, to a point where private school attendance became a disadvantage in some instances. It is possible that the advantage initially found was due to school compositional effects – children from a low socio-economic background benefited from mixing with children from a high socio-economic background (Wylie, 1998). Toma's analysis showed that the association between socio-economic mix and mathematics achievement was larger than it was between school type and mathematics achievement. Wylie (1998) cautions that Toma's findings may not be representative of all New Zealand private schools, since between half and two-thirds of her private schools were integrated Catholic schools; as such, the effects might be attributable to the ethos and organisation of Catholic schools (cf. Bryk et al., 1993) and not to private schools *per se*.

Recently, a scheme called targeted individual entitlement (TIE) scheme, offering scholarships for private school places to children from low-income homes, began in New Zealand. The scheme, based on the assumption that there is an advantage for students attending a private school, tended to attract parents who were better educated than their low-income peers, more likely to be in skilled work, and slightly more likely to have attended a private school themselves (Wylie, 1998). Similarly, the ethnic profile of selected students was not representative of the ethnic composition of low-income families. An evaluation of this scheme indicates that TIE students are by and large perceived by teachers and principals to be doing as well as or better than fee-paying students (Gaffney & Smith, 1999). Over 90% of families involved in the scheme reported they were satisfied with the social and academic progress their children were making and reported higher levels of achievement for their children. Although reports are very positive, it is unclear whether students' achievement levels are higher than they would have been as a result of attending a public school.

With respect to single-sex versus coeducational schools, the findings of New Zealand studies parallel those of overseas studies. Early research in New Zealand found that students viewed single-sex schools as more academically oriented and satisfying than coeducational schools (Jones, Shallcrass, & Dennis, 1972). More recently, Nash and Harker (1998) examined 12 single-sex schools and 25 coeducational schools, focusing on the attainment of girls, using data collected as part of their Progress at School project. HLM analysis techniques revealed no significant differences in the School Certificate performance of students attending single-sex and coeducational schools. Using national data, Nash and Harker also compared the number of Bursary awards obtained and the number of students retained to Year 12 in the two types of schools. Again, they found little relationship between the percentage of students obtaining a Bursary award and the type of school they attended. These findings are consistent with those of recent United States studies, showing no clear evidence of differences in achievement between students attending single-sex and coeducational schools, for either boys or girls.

We know of no New Zealand research comparing the performance of home-schooled students to students in public or private schools. However, the Home Schooling Federation of New Zealand (1996) cites anecdotal evidence that seems to confirm that home-schooled pupils exhibit superior educational performance to students in public or private schools. Similarly, New Zealand home-schooling parents who responded to an exploratory study conducted by Kerslake, Murrow, and Lange (1998) reported that they perceived home schooling improved their children's academic progress. The parents also reported that social needs of their children were met by joining with other home-schooling families and participating in sports and church groups. These results cannot be taken as representative of all New Zealand home schooling families, as the survey had a low response rate, but they are consistent with findings of recent research in the United States.

5.3 School Size

School size is a factor that could moderate effects of school composition, as the size of a school may influence student learning over and above individual characteristics students bring to a school. Research examining the effects of school size is limited, as the common-sense view that school size relates positively to achievement has only recently been questioned (Howley, 1995).

There are two important and opposing viewpoints evident in school-size research (Lee, Bryk, & Smith, 1993). One viewpoint is based on an economic efficiency argument in favour of 'economies of scale', which claims that financial savings accrue when costs are spread over larger numbers of students. It is also suggested that larger schools, because they have greater numbers of students with similar needs, are better placed to create specialised services to meet these needs. In contrast, smaller schools are forced to focus their resources on core programmes. Evidence that increased school size equals greater resource strength is contradictory, however, and it is unclear whether the cost benefits claimed for large schools ever materialise (Lee et al., 1993). Some studies find that, as the number of students served by a school or district increases, more fiscal resources become available for teachers' salaries, instructional materials, and support for professional development. However, the academic consequences of economies of scale and greater resource strength are not clear (Lee et al., 1993).

The other viewpoint apparent in school-size research is concerned with the influence of size on the formalisation of social interactions and the consequences that flow from this formalisation (Lee et al., 1993). Advocates of small schools claim that smaller schools, particularly small high schools, can maximise interrelations among students (Carnegie Foundation for the Advancement of Teaching, 1992), and research findings offer support for this view (Bryk & Driscoll, 1988; Lee & Smith, 1995, 1997).

Although empirical research has been neither numerous nor strong, several noteworthy United States studies have recently examined the effects of school size. These studies focus on high school rather than primary school and generally find a negative relationship between size and achievement, consistent with the view that small schools can maximise interrelations among students (Howley, 1995; Lee & Smith, 1995, 1997). They also find that small schools provide a more equitable learning environment (Howley, 1995; Lee & Smith, 1995).

Lee and Smith (1995) examined the effects of school size on student achievement and the social distribution of academic gains. Small size was considered a feature of school structure that moved schools towards a communal organisation. Using the NELS 1988 data, Lee and Smith found that students in small schools learned more in reading, maths, history, and science. The effects of increased school size on cognitive gains were negative and significant, with effect sizes ranging from –.30 to –.40. Furthermore, students in small schools were more engaged in their courses. Achievement was also distributed more equitably in smaller schools. Magnitudes associated with social equity (as defined by slope coefficients for the relationship between academic achievement and socio-economic status) ranged from small (effect size = .03) for maths to moderate (effect size = .34) for reading. Lee and Smith suggest that school size has an indirect effect on learning and engagement as it can affect the economic, academic, or social organisation of high schools, and these characteristics could in turn have consequences for students.

Howley (1995) examined the relationship between school and district size, socio-economic status, and achievement in a recent United States study. West Virginian schools and school districts were used as the unit of analysis. School size was defined as the number of students in each district enrolled in each school Year under analysis (Years 4, 7,10, and 12). Howley examined the role that socio-economic status played in regulating the effects of increasing school size on student achievement. Socio-economic status was defined as the proportion of students in each school receiving free or reduced-cost lunch. Analyses revealed an interaction between school size, socio-economic status, and achievement for three of the four Year levels. Moreover, the magnitude of the effects increased with Year level. Effect sizes ranged from –.01 in schools where five percent of students received free or reduced-cost lunch (a measure of high socio-economic status) to –.11 in schools where 95% of students received free or reduced-cost lunch (a measure of low socio-economic status) at Year 7. At Year 10, effect sizes ranged from –.03 at the five-percent free or reduced-cost lunch level to –.56 at the 95% free or reduced-cost lunch level at; and, at Year 12, effect sizes ranged from .28 at the five-percent free or reduced-cost lunch level to –.69 at the 95% free or reduced-cost lunch level. The results of Howley's study parallel those of Friedkin and Nocochea (1988) in showing that large schools may benefit high socio-economic status students but that small schools may benefit low socio-economic status students. Large schools magnified the disadvantage of low socio-economic status students and small schools reduced the advantages experienced by high socio-economic status students.

With the aim of identifying the optimal size for a secondary school, Lee and Smith (1997) examined how growth in reading and mathematics achievement was influenced by secondary school size. They also examined the relationship between secondary school size and the equitable distribution of learning. Students who attended high schools that enrolled between 600 and 900 students were found to have optimal learning. Size effects were larger for learning mathematics (over one standard deviation in some cases) than for reading, where gains were moderate (between 0.3 and 0.5 of a standard deviation). However, a similar pattern was found for both subjects. Learning was also found to be more equitable in schools enrolling between 600 and 900 students. Differences in achievement between high and low socio-economic status students in schools in the 600 to 900 enrolment range were about two points on a 40-point test. In schools with fewer than 300 students, differences were about 3.5 points, and, in the largest schools, the differences were about five points. Of concern is Lee and Smith's finding that the average socio-economic status in very large or very small schools was low, as this suggests that large numbers of socially disadvantaged students attended high schools of a size where they are likely to learn the least.

Ideal school size does not vary according to the characteristics of the students who attend. Schools in the moderate size range (between 600 and 900 students) produce the greatest achievement gains for high and low socio-economic status students and for schools with high and low proportions of ethnic minority students. Although the results of Lee and Smith's study suggest there is an optimal school size, they suggest that size acts as a facilitating or debilitating factor for other aspects of school organisation and, in this indirect way, influences student learning. Rather than building smaller high schools, Lee and Smith recommend creating 'schools within schools', inside larger high schools. To avoid further specialisation and social stratification, they also recommend the units be representative of the demographic make-up of the school as a whole (for an overview of various approaches to organising schools into small units to enhance educational achievement and equity, see Oxley, 1994).

In summary, economies of scale arguments, which suggest that students in large schools are at an advantage academically, are not substantiated by research. However, there is considerable support for the view that small schools can optimise interrelations among students, and there is good evidence to suggest that smaller schools enhance academic achievement, particularly the achievement of disadvantaged students. It is important to note that schools can be too small, to a point where the ability of the school to match the curriculum delivery of larger schools is undermined because of 'diseconomies of scale'. Rather than benefits on learning and engagement being a direct result of school size, Lee and Smith (1995) state that the size of a school can affect the economic, academic, or social organisation of high schools and that these characteristics can in turn have consequences for students. Research suggests that small schools are more likely to have positive social environments and less diversification of instruction. Findings that disadvantages experienced by low socio-economic status students are compounded over time (as are the advantages experienced by high socio-economic status students) are concerning. Also worrying are findings that disadvantaged students are more likely to be in schools (very large or very small) where they are least likely to learn. However, moderately-sized schools appear to disrupt these patterns.

We know of no New Zealand research examining effects of school size. However, there are frequent observations that students in small schools, particularly in rural areas, achieve at a lower level than their urban counterparts, and this may be a reflection on the socio-economic and ability intake of these schools (Ministry of Education, 1991). A report of the Education Review Office (1999) on small schools concluded that, in almost all areas of curriculum management and delivery, small schools (those enrolling between 50 and 150 students), had lower levels of performance than a group of larger schools. The report was based on 500 performance reports on full (Years 0 to 8) or contributing (Years 0 to 6) state primary schools, 400 of which were classified as small schools. According to the report, small schools performed well on indicators of school climate and of relationships within the school (between the board, principal, and staff, and between staff and students) and of relationships between the school and the community. These findings are broadly consistent with those of the recent United States studies that smaller schools promote positive interrelations among students and staff.

5.4 Implicating Peer Effects on Learning

A common theme of the earlier sections of this chapter is that studies of school compositional effects say little or nothing about the possible range or nature of processes underlying the phenomenon. Willms (1992) provides as much detail as any researcher in this area when he comments:

> *Schools with high social class or high ability intakes have some advantages associated with their context: on average they are more likely to have greater support from parents, fewer disciplinary problems, and an atmosphere conducive to learning. They are more likely to attract and retain talented and motivated teachers. Also there are peer effects that occur when bright and motivated pupils work together (p. 41).*

Yet, this level of discussion can in turn be compared with a study by Thrupp (1999b), which focuses explicitly on the mechanisms that might underpin the effects of school composition (at least those associated with 'school mix', the socio-economic composition of the student intake). This study used a mostly ethnographic approach to explore the possible impact of school mix on various school processes, including reference group processes, classroom instruction, and school organisation and management in four New Zealand high schools of varying socio-economic composition. Thrupp argued that school mix appears to impact on school processes in numerous ways so as to cumulatively drag down the academic performance of schools in low socio-economic settings and boost it in middle-class settings. Conversely, by virtue of their higher socio-economic status, the matched students attending the middle-class schools in the study ('Victoria', 'Wakefield', and 'Plimmer' Colleges) accrued numerous advantages over those in the lowest socio-economic status school ('Tui College').

What might explain these effects? To begin with, in the middle-class schools, the matched students' friends and classmates had a wider range of curriculum-relevant experiences, higher levels of prior attainment, more previous experience of school success, more regular school attendance, higher academic goals, higher occupational aspirations, and less involvement in 'alienated' student subcultures than those at Tui College. When it came to classroom instruction, the matched students in the middle-class schools were taught in teaching classes that were generally more compliant and more able to cope with difficult work. They used more demanding texts and other teaching resources and their teachers were more qualified and more motivated. The middle-class schools were also able to support more academic school programmes and a wider range of extracurricular activities. Finally, it was easier to organise and manage the middle-class schools. Day-to-day routines were more efficient and more easily accomplished. They had less pressured guidance and discipline systems, with higher levels of student compliance and fewer very difficult guidance or discipline cases. Their senior management teams had fewer student, staff, marketing, and fund-raising problems, and had more time to devote to planning and to monitoring performance. The middle-class schools also had boards of trustees with more useful qualifications and business contacts.

The study suggests that these differences were only partly to do with disparities in material and staffing resources – factors often cited as a cause of inequalities between schools. A more powerful explanation revolves around the idea of school policies and practices of many kinds having to be negotiated with students on the basis of class-related levels of compliance, motivation, and ability, which are, in turn, related to students' views of schooling and their likely occupational futures. The study also points to the importance of critical mass. In a predominantly middle-class school, the struggles of working-class families and students are marginalised and can have relatively little effect on school management, teaching, and student reference group processes. As a school becomes more working class, however, the processes of the school will shift, despite resistance from middle-class teachers and students, towards the culture of the increasingly sizeable working-class group. This in turn reflects the organic or interconnected relationship between schools and middle-class, rather than working-class, families (Bourdieu & Paseron, 1977). Schools develop processes that reflect their socio-economic mix. Solidly middle-class schools have strongly supportive student cultures that allow the schools to teach an academic, exam-based curriculum and to organise and manage themselves relatively smoothly. Working-class schools are, in general, quite the opposite. Yet Thrupp stresses that a deficit approach to working-class culture is not intended here. Rather, the study is seen to highlight the extent to which school effectiveness in an academic sense not only reflects the middle-class bias of schooling in capitalist societies, but also appears to rest upon the cultural resources and responses of students from middle-class families.

There are clear parallels between the peer processes reported here and those reported by other ethnographic studies that, although investigating somewhat different questions, have also compared schools of differing social class. Of particular relevance is research by Anyon (1981), Lareau (1989), and Metz (1990) in the United States, research by Connell et al. (1982) in Australia, and research by Brown, Riddell, and Duffield (1996) in Scotland. There are also parallels between Thrupp's emphasis on the importance of 'critical mass' with the analysis presented by Harris (1998), who points to students adopting the behavioural norms of the majority of students in their schools.

In view of the strength of the work showing possible benefits of Catholic private schools, at least in the United States (Bryk et al., 1993; Evans & Schwab, 1993a, 1993b), it is worthwhile examining attributes identified as unique to these schools that might implicate peer effects on learning. Bryk, Lee, and Holland's (1993) findings suggest that the differential pattern of school effects for public and Catholic private schools work primarily through features of school organisation. Although direct links between school organisation and student achievement are difficult to make, as we have already mentioned, Bryk and colleagues suggest that schools organised as communities have direct consequences for teachers and students alike. They suggest that even though the effects of communal organisation begin with the teachers, communal organisation has powerful social consequences for students. In their words:

> *The presence of highly committed teachers is likely to be infectious. Drawing faculty together results in a social solidarity that also draws students into the mainstream of school life. The actual processes through which this occurs are probably quite complex: the personal interest of individual teachers in individual students fosters a social bonding of these students to the school and to the core activities that manifest the school's goals. When the social activity is widespread, a normative environment is created in which caring and a sense of hope and purpose come to characterise the personal experiences of both adults and students (p. 276).*

Although their research does not allow us to make direct links between communal organisation and peer influences on student learning, a communal organisation and inspirational theology may enable Catholic schools to harness aspects of the ambient environment to their advantage. As students internalise the values and norms critical to a communally organised school, it is possible that ambient influences come to contribute to learning from peers. Many ambient mechanisms have their effect on proximal indicators of learning, such as academic motivation, rather than direct effects on achievement. This is consistent with Bryk, Lee, and Holland's (1993) hypothesis that the major effects of communal organisation are located in the personal and social domains rather than in the academic domain.

Bryk, Lee, and Holland (1993) also emphasise the effect of an inspirational ideology, another key factor they identify as unique to Catholic schools. It is possible that part of the reason Catholic schools are successful can be attributable to their ability to utilise an inspirational ideology to establish a set of social norms (an ambient mechanism). Bryk, Lee, and Holland acknowledge that there are no variables in the HSB data that measure 'inspirational theology', and they note that the influence of this concept is likely to be complex. They claim its impact is apparent in aspects of school organisation such as "the use of conventional instruments, such as tracking while consciously avoiding the reproduction of inequity" (p. 303). Although ideology is difficult to measure, it should not be ignored, as it may be more pervasive than other characteristics of Catholic schools, such as academic organisation or communal school structure.

5.5 Conclusion

We began this chapter by pointing out that the assorted quantitative literature on school compositional effects offers inconsistent findings. Taking the stance, for the purposes of this review, that a range of conceptual and statistical issues may explain these inconsistencies, we have considered the literature in relation to an ideal model. This analysis suggests the existence of a school compositional effect when judged against the criteria set out earlier in this chapter, although we also found only one study (Bryk, Lee & Holland, 1993) which sought to explore the full school compositional effect hypothesis. As we saw, the New Zealand school effects literature also reflects similar inconsistencies and problems.

There are likely to be other, more deep-seated explanations for the inconsistencies than can be explained by different combinations of statistical methodology, explanatory variables, and outcome measures. Something important seems to be missing in the conception of how and when school composition might influence student achievement. As indicated earlier, we think the key problem is a lack of concern with theorising about what might be involved in creating school effects (i.e., the likely underpinning mechanisms and mediating factors). This is because studies of contextual and compositional influences on educational outcomes have used an input-output formulation and have, crucially, omitted to model how these influences act through the processes of schooling. This problem was pointed out some time ago by Erbring and Young (1979), who demonstrated the fallacy of ignoring the nature of the social structure between peers through which variables such as ability and socio-economic status probably act. For example, assuming that a variable such as the social composition of a class has a direct influence on the outcomes of the class members is of little value if the exact mechanism and necessary conditions are not specified in the model. This holds for both the regression models available at the time Erbring and Young were writing and the multilevel (e.g., HLM) models since developed.

The fact is that many studies may not have entertained the degree of complexity involved in understanding school compositional effects. We are dealing with weak statistical evidence, both for and against these effects. It is quite possible that none of the statistical analyses that have been carried out to date have truly captured the effects of school composition. For example, there is at least one study, Fitz-Gibbon (1997), using a consistent overall methodology that has found positive and negative effects as well as an absence of effects with the same variable at different stages of schooling (Thrupp, 1999b).

There is also the problem that research, especially research with important policy implications, is never neutral but coloured by the politics of its time. In the case of the literature on school composition, the influence of the school effectiveness movement on the findings of researchers should not be underestimated (Thrupp, 1999b). It should also be remembered that many studies have occurred in the context of advocacy for particular policy positions. The heavily criticised work of choice advocates, such as Chubb and Moe (1990), is perhaps the most obvious example. It also seems likely to us that the political malleability of school compositional research is not unrelated to the problem of adequately modelling school compositional effects. For instance it may that it is the inability of researchers to properly capture school-level compositional effects that has allowed their findings to be more easily swayed by politics.

Turning to the question of how school compositional effects might be related to peer effects, we again strike the problem that considerably more research energy has gone into trying to measure school compositional effects than into conceptualising them (i.e. trying to understand what they might represent, how they might work, what factors might mediate their effects, and so on). Thrupp's (1999b) study focuses explicitly on the mechanisms that might underpin the effects of school composition and goes some way towards indicating how school compositional effects may be related to compositional effects at other levels. It also suggests the possibility of direct effects of school composition on learning among peers, through the way school composition may influence school organisational and management processes.

With respect to particular dimensions of school composition, it is difficult to establish the extent to which advantages of schools at the 'extremes' of different student compositions (e.g., public versus private) truly reflect advantages due to the schools' composition. In these circumstances, school composition is inextricably interwoven with school resources, parental values, and students' background characteristics. There does, however, appear to be strong evidence for positive effects of Catholic private schools, at least in the United States, on learning outcomes. Furthermore, Catholic private schools seem to provide a more equitable learning environment for students, with increased learning for low-ability students as well as advantages for high-ability students. It is possible that these benefits are the result of a communal organisational structure and inspirational ideology. Positive effects, especially for low-ability students, may also be attributable to a common core curriculum and higher academic focus in Catholic schools in the United States.

There is also good evidence that school size is a mediating factor in effects of school composition. Students appear to benefit from attending smaller high schools, and smaller high schools appear to provide a more equitable learning environment. This is attributed to more personalised and intimate social relations. Small schools, because of their limited resources compared to larger schools, are more likely to focus these resources on the provision of a core curriculum, whereas larger schools tend to diversify to meet individual student requirements. This focus on a core curriculum may also benefit students. However, school size is not considered to have a direct effect on achievement; rather, it influences the economic, academic, and social organisation of the school, and this in turn may affect learning.

5.6 Recommendations for Further Research

- There is need for more research into the effects of school composition that incorporates all features of the 'ideal' model of large-scale statistical research outlined in this chapter.
- Statistical models of school compositional effects provide a relatively 'blunt instrument' when it comes to understanding the process by which any effects of school composition arise. Studies employing these techniques need to be complemented by development of theory that explains how school composition effects come about and by more fine-grained descriptive or ethnographic accounts, similar to that of Thrupp (1999b), that can detect the presumbly subtle processes that underlie such effects.
- Based on this theoretical development, school-effect studies in New Zealand need to be based on larger samples of schools that are representative of the country as a whole, and they need to use highly reliable, standardised outcome measures that provide data for the entire age cohort in the sample.
- It may be useful to investigate the ethos and organisation of Catholic private and integrated schools in New Zealand to see if there are parallels with findings from United States research.
- There is need for research that examines the implications of school size in New Zealand, addressing questions related to the social distribution of learning and to determine whether trends mirror those of overseas research. In particular, this research should consider how communal organisation influences learning outcomes in small and large schools in New Zealand.

CHAPTER 6

MODELLING PEER EFFECTS

In the previous chapters, we have sought to identify compositional effects of various groupings of students defined on the basis of student ability, socio-economic status, gender, and ethnicity. We have described effects for groups of students in particular curriculum areas (e.g., homogeneous ability groups in reading or maths), and we have examined special populations of students (e.g., special needs or gifted students) in so far as these students have been included in these groupings. Where we have found evidence of compositional effects, we have sought empirical and theoretical support for the extent to which they might implicate peers. We defined peer effects as the influences of student-to-student interactions and group dynamics on learning outcomes, where the term 'group' is used in the sense of any aggregation of students, be it a pair, small group, whole class, or school.

In this chapter, we develop a conceptual model of peer influences on learning outcomes. We lay the foundations for the model by first identifying issues that are common across the assorted literature and possible linkages across the domains that may point to the mechanisms and key channels through which peers influence learning. We then describe our conceptual model. These sections highlight possible disparities between findings from outcome-based studies of student learning and findings from descriptive studies that do not include measures of learning outcomes. We then present explanations that might account for the inconsistencies in findings. We conclude by highlighting policy implications of our conceptual model and we point out areas requiring further analysis.

Our model attempts to account for the empirical and theoretical foundations derived from our review of the literature. It lays out our best judgement as to the relative magnitudes and direction of compositional and peer effects, as well as those associated with family and school resources, and it incorporates our informed vision as to the environments in which peer learning occurs and the mechanisms and processes that mediate learning among peers.

6.1 Foundations

We have identified ambient environments that foster learning among peers organised into groups, classrooms, and schools, and a continuum of tutorially configured environments that foster learning, ranging from learning that happens to occur in a social context (e.g., peer tutoring) to 'true' socially constructed learning (e.g., collaborative knowledge construction). Learning is directly and indirectly influenced by peers, though we note that learning is also influenced by teachers, since they choose various configurations to structure the learning environments – be they pairs, small groups, or whole classes. Within these environments, we have identified various mechanisms and processes that mediate learning among peers.

Our review has also revealed some commonality across the literature in the manner in which educational structures impact on learning outcomes. The theoretical stance we have taken at the outset of this review, following the work of Barr and Dreeben (1983; Barr & Dreeben, 1991), is that educational structures have only indirect effects on student learning. The results from our review, at all levels of aggregation, are that compositional effects measured in terms effect sizes are small and show substantial variability, particularly at the school level. Hence, theory and data converge on the notion that the effects of educational structures are indirect and probabilistic in nature. That is, the effects of these structures (e.g., streaming, class size, school mix) are mediated by an array of instructional and peer processes, and the presence or otherwise of the structures can change the probability that these processes occur, which then influences student learning. So, for example, reducing class size does not **directly** influence student learning. Rather, reducing class size merely increases the probability that the environment can be structured to capitalise on peer influences such as changing self-efficacy, enhancing academic reputations, altering expectancies for success, and spreading contagion effects. Indeed, we noted many instances where changing these class structure led to no change in the manner in which teachers configured interactions, no change in the nature of the curricula and instructional strategies used by teachers, and no change in the interactions between students. Hence, our claim is that the school, class, and group compositional effects, at best, change the probabilities that successful learning conditions can be constructed. These changes can then lead to positive impacts on student learning.

Moreover, we have discerned a pattern to the compositional effects from our review. Not only are average effect sizes small – suggesting our indirect and probabilistic notions – they change in levels of magnitude from the school to the class to the small instructional group. The effects of these structures on student learning are relatively larger at the group level, somewhat smaller at the class level, and smaller still at the school level. Specifically, we estimate that the effects we have observed for within-class grouping average about .25; for class-based configurations, they average about .10; and for school-level influences, we judge they average about .05 (although there is considerable variability at the school level). If this pattern of effects is correct, then it lends credence to the notion of compositional effects operating through a series of nested hierarchical layers; effects are greatest at the 'coalface' where learning occurs and become smaller at the more distant layers. This is not to deny that there may be direct effects of school and class on individual learning; there may be, but we believe they are small and variable.

These effects range from tiny to moderate. In Chapter 1, we detailed effect sizes based on over 400,000 studies of educational characteristics and reported an overall mean effect of .40. There were few characteristics relating to achievement outcomes as low as .10, so the effects of school composition and class-level configurations are among the lowest influences. Within-class grouping effects are still relatively small, but there are worthwhile patterns of effects at this level that are worth exploring. The effects for many peer learning environments, mechanisms, and processes are potentially above average, though we have not specified effect sizes for each of these (much of the literature on these topics does not lend itself to meta-analyses). We would encourage further detailed analyses of their effects on learning outcomes. The overall effects of school composition, assuming an effect size of .05, is akin to a success ratio of two percent; of between-class influences, assuming an effect of .10, is akin to a success ratio of five percent; of within-class groupings, assuming an effect of .25, is akin to a success ratio of 12%; and the various mechanisms perhaps 10-35% (see p.7 for definition of success ratios). It is crucial to note, however, that there can be much variability in the effect sizes for specific compositional effects, for specific student sub-populations, and within some schools. These estimates are overall findings, and the specifics, detailed in the preceding chapters, are critical.

One feature that seems common at all levels of analysis is the notion of reciprocity of influences between students, and teachers, and school organisation and management. At the group level, reciprocal student-teacher influences seem to establish group cultures or norms of behaviour that might differentially support learning. At the class level, different streams or groups of students may actively construct different classroom cultures, as their actions and reactions influence what is taught – what Thrupp (1999b) refers to as the 'negotiated curriculum'. At the school level, different classroom environments may constrain (or enable) different organisational and management processes. For example, a powerful, though variable, reciprocal effect occurs when the relative abilities of students within a class lead teachers to make curriculum and teaching decisions related to the cohort of students in their class rather than to some (high) standard of performance established by the curriculum (Timperley, Robinson, & Bullard, 1999). Similarly, it may be that the perceived or real perception of within-class effects can lead parents to choose other seemingly more advantaged schools. For example, a weak reciprocal effect from the class to the school level may occur in situations of 'bright flight', where the exit of students results in a change in the school mix. We find, however, that these reciprocal effects from the student level to group, class, and school are weak, variable, and not nearly as powerful as other effects.

We determine that there are two major sets of resources that impinge on a model of peer learning. The first set relates to family resources and consists of parental encouragement and expectations, and the other family resources (e.g., home language and ethnicity) that are brought to bear on students' experiences with school work. The second set of resources relates to curriculum and teaching resources and includes the richness and appropriate challenge of curricula for students, and the proficiencies of teachers in configuring learning experiences for their students. We deliberately exclude financial considerations from this model, as these appear to be relatively minor in terms of their influence on peer effects and are the subject of another Ministry of Education review.

Our review indicates that the effects of family resources are relatively larger at the school level and decrease at the class, then group levels, and have minor influence at the level of peer learning environments. We have two reasons for making this claim. First, peer effects seem to get larger the closer we are to the contexts in which they operate, as we have already observed. Second, by definition, as soon as we move to the class or the group as the unit of analysis, we are implicitly partialling out the effects of family resources, because there is less variance in these factors within classes and within groups. Hence, between-school variance in outcomes that is associated with family resources should be large, whereas between-class variance associated with these factors should be smaller (particularly in homogeneous school populations), and between-group variance should be smaller still (particularly in streamed classes). Variance between students within formal and informal groups that is associated with family resources should be of little consequence, relative to other factors; indeed, the basis on which teachers most frequently select students (or students self select) for these groups tends to mitigate against differences in family background. Even in groups of mixed ethnic composition, where ethnicity could come to the fore to influence students' interaction and learning, teachers play a powerful role in establishing expectations for relative performance.

The opposite is the case for the curriculum and teaching resources, although the evidence is sparse. Curriculum and teaching resources have influences at the school level in terms of the stated curriculum and enunciated values promoted by the school. These influences often lead to some parents 'deciding' to send their children to particular schools, although the norm in New Zealand is still a preference for the neighbourhood school (whatever its enunciated values). The intention of many class-level configurations is primarily related to curriculum and teacher factors, although our review has highlighted that this intention is rarely realised in practice. Teachers do change their instructional methods according to class composition, but more often the change is one of pace (slower or faster) and not one of method. Too often, the curriculum is covered to a greater or lesser extent according to class composition, and too infrequently are richer and different experiences offered to differing groups to cover a similar curriculum. At the within-class group level, a primary purpose is to optimise the interaction between teaching methods and curriculum opportunities, and the evidence is that this is more likely to occur at the within-class level. Certainly, within the various tutorially configured interactions, the primary purpose is exactly to alter curriculum and teaching methods; hence the influence is greatest at this level.

Further, it appears that peers have more influence when they have 'knowledge' or understandings to share or seek and when they are placed in tutorially configured situations that foster the seeking of collaborative understandings. These understandings are more likely to occur when the curriculum tasks are appropriately constructed, the goals of the tasks are more specific than vague, and there is a sense of shared commitment to achieving these goals. We note that many of the mechanisms and processes of learning can then be invoked to optimise peer influences on the tasks. This implies that the role of the teacher is fundamental in configuring such environments and providing structures that optimise peer influences. There may be instances in which these structures are enhanced at the class and school level, but they are less powerful.

6.2 Conceptual Model

Taking these empirical and conceptual foundations into account, we propose a multi-layered model with effects propagating from school-level influences to class-level influences to group-level influences to ambient and configured environments for learning among peers. We present this model in Figure 6.1. The three planes at the school, class, and group levels represent all manners of influences on learning associated with the respective levels. The vertical arrows represent the compositional, including peer, effects propagating from the different levels. The arrows in our model are not intended to be viewed as causal paths, but should be considered as arrows of probable influence; that is, they indicate the most probable direction of influence on student learning. We propose that the bulk of the effects are indirect. Hence, peer effects 'look' smaller the further we move away from the instructional coalface, because they are mediated by intervening layers and because teachers bear primary responsibility for shaping students' learning experiences. We leave open the possibility that there may be direct effects on peer learning environments, shown by dashed arrows, but we believe such effects are small.

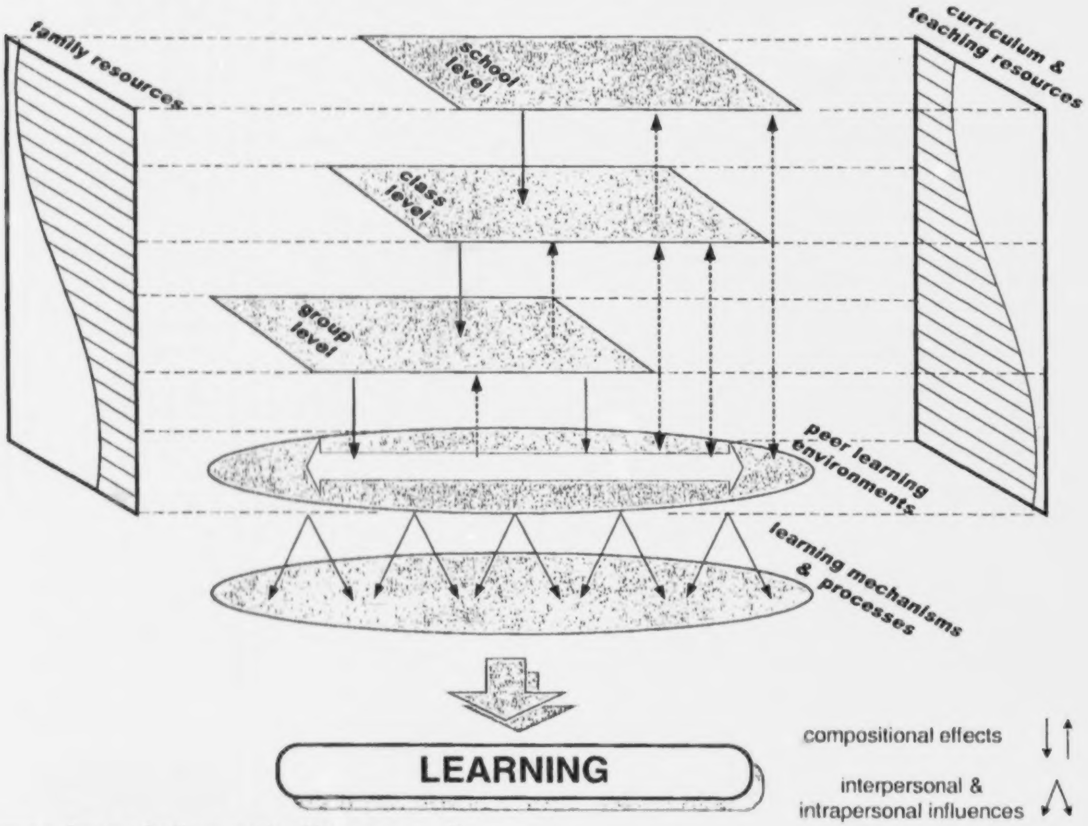

family resources
school level
class level
group level
curriculum & teaching resources
peer learning environments
learning mechanisms & processes
LEARNING
compositional effects
interpersonal & intrapersonal influences

Figure 6.1 Conceptual Model of Peer Influences on Learning

In concert with the three planes of influence on learning, we argue that family resources have greater effects at upper layers and smaller effects at lower layers, though the effects never completely disappear. Conversely, curriculum and teaching resources have greater effects at lower layers and smaller effects at upper layers, though, again, they never completely disappear. The relative magnitude of effects of family resources, curriculum and teaching resources, and peers on learning are difficult to estimate precisely, although we are convinced that the effects of compositional factors are minor. Home and school supports for learning carry the lion's share of the weight in predicting student outcomes (Lee & Croninger, 1994).

It is important to note that the effects between layers of school organisation are, by definition, multiplicative. In other words, the total indirect effects are computed by multiplying together the direct effects of each layer on the next. Hence, when the effects are multiplied to yield the total indirect effect on student learning outcomes, the size of these effects is small at the group level, smaller still at the class level, and very small indeed at the school level.

The direct effects of each layer of organisation on the next have rarely been estimated, nor do we know much about how compositional effects at higher layers (e.g., school) condition compositions at lower layers (class or within-class instructional groups). Two empirically based descriptions of these 'compositional transformations' provide insight into the processes. Garet and Delany (1988), in a study of four high schools in the United States, showed that school composition can affect students' course-taking patterns (i.e., their placement in classes for particular subjects). After controlling for characteristics of individual students, those students attending schools with low average ability had a greater chance of taking advanced maths classes than did similar students in schools of high average ability – because of the need to fill available places in the advanced classes in the low-achieving schools. Similarly, Dreeben and Barr (1988), in a study of seven schools and 13 Year 2 classes in the United States, showed that class composition constrained the composition of instructional groups and, hence, the nature of the instruction and students' learning. In 'difficult' classes, defined as large classes with low mean aptitude and a large number of low-aptitude students, teachers divided students into groups of roughly equal size that were thus 'bottom heavy' (i.e., there were too many students in the low-aptitude groups). By contrast, in 'easy' classes, teachers divided students into groups of unequal size with small low-aptitude groups. As a result, groups of comparable ability learned less in the 'difficult' classes than in the 'easy' classes. These accounts illustrate how student composition at each layer of organisation shapes the arrangement of students, and the nature of instruction, at lower layers.

The upward arrows in Figure 6.1 show the reciprocal influences of peers. These capture the notion that peers may also be actively involved in influencing teachers and school organisation and management. We are not sure of the magnitude of these reciprocal influences; findings from descriptive studies, particularly those from sociolinguistic and ethnographic studies, suggest they are real, but the outcome-based studies shed no light on their sizes or effects.

Compositional effects can influence student learning by affecting the ambient and tutorially configured environments that mediate learning among peers. It appears that the school-level composition is more likely to affect the ambient environment (particularly helping, friendship, and peers as models) than the tutorially configured environment. Similarly, the between-class composition is more likely to affect the ambient environment, whereas within-class grouping is more likely to affect both ambient environment and tutorially configured interactions. This may be particularly true at secondary school, in specialised subjects, where student motivation, interest, volition, and aspirations may be more important determiners of progress than they are in primary school and where teachers are less likely to form students into small instructional groups (cf. Dreeben & Barr, 1988). For this reason, we have allowed the possibility that class-level composition may impact directly on peer learning environments (mostly ambient) or be mediated by group-level configurations, depending on whether schooling is at the primary or secondary level. Stronger reciprocal influences probably stem from ambient environments, as the teacher is more likely to change the group or class configuration depending on the normative environment among peers (although the more effective teachers probably spend much time creating climates rather than reacting to them).

The model depicts the mechanisms and processes that mediate learning as the final layer. These processes include social comparisons, socio-emotional support, feedback, cognitive restructuring, rehearsal, internalisation, and activating inert knowledge. The model carefully specifies that there is no one-to-one mapping between peer learning environments and these mechanisms and processes. Learning is multiply determined. It is conceivable that various mechanisms and processes, depicted in this final layer, may enhance the likelihood that certain peer learning environments will be used. For example, social norms favouring cooperation may foster helping behaviour, peer tutoring, and even cooperative learning. Hence, there may also be reciprocal influences between mechanisms and processes and peer learning environments. However, we do not know of any empirical literature documenting these reciprocal influences so we have erred on the side of caution and not depicted these in our model.

Within these layers, different orders of homogeneity/heterogeneity of student composition (defined in terms of ability or socio-economic status or, perhaps, gender and ethnicity) impact differently on ambient and configured learning environments. In situations where students are relatively homogeneous, educational structures implicate the probability that peer influences on learning are mediated more by ambient mechanisms (e.g., reference group processes, norms of behaviour, or institutional effects) than configured mechanisms (cf. Harris, 1995; Thrupp, 1999b). For example, in configurations comprising students who are homogeneous in ability, we consider that peer group norms may be a key mechanism that mediates learning. We caution, though, that these norms probably arise from cycles of reciprocal student and teacher influences that evolve over time and into which participants are socialised. Hence, they may be indistinguishable from what is traditionally thought of as school or classroom 'climate'. Moreover, because peers are intimately involved in this socialisation process, we have construed this as a 'peer effect' – yet teacher expectancies and instructional processes are also involved. Theory and logic suggest that it does not make much sense to try to separate these factors, or to identify the relative contributions of peers and teachers, despite the fact that these dynamics do not represent a pure peer effect.

Relatively homogeneous groupings of students constitute the bulk of the educational structures we have examined in this report. Certainly at the school and class levels, and somewhat less at the within-class level, most students are grouped homogeneously – for example, by ability or gender or by reducing class size (hence reducing student variability). The constant claim by teachers and principals, and often by parents, is that this homogeneity is more conducive to instructional practices. This may be so, but there is less evidence that such compositional effects translate into learning outcomes. Our model specifies that the major effects of these grouping practices are to some extent related to the ambient environments and more specifically to the tutorially configured interactions, as they implicate both mechanisms and processes. More relevant, the review has identified that, too often, configurations at higher layers of the model do not lead to differential instructional behaviours. Smaller class sizes, streaming, single-sex classes, within-class grouping, and so on do not guarantee that curricula will be modified (other than in pacing) or that the opportunities claimed for homogeneity will be realised in adapting interactions and mechanisms to the cohort of students. It is suggested, particularly in light of the evidence of accelerated schooling, that there may be similar processes and mechanisms that are effective for most students, and that the art is to construct the tutorially configured interactions and ambient environment to optimise these, within the class. More important, it is recognised that the successful implementation and use of these influences is more a function of the teacher than the composition of the class. We note the particularly powerful effects of norms of behaviour, both by the teacher and by the students, as mediators of these influences.

By contrast, in situations where students are relatively heterogeneous, educational structures implicate the probability of peer influences on learning coming more from tutorially configured mechanisms (e.g., cooperative and collaborative learning). For example, in structures comprising groups of students that are heterogeneous in ability, particularly where the range of abilities within the group is not too wide, the cooperative interactions identified by Webb (1991) probably play a role. In these situations, giving and receiving elaborated explanations and feedback that is timely and responsive to students' needs is a key causal mechanism by which peers influence learning. Receiving timely, relevant, and elaborated help may enable students to correct their misconceptions and may foster greater engagement and constructive, problem-solving activity. Low-ability students derive more benefit in heterogeneous groups because they are more likely to receive timely and elaborated help from their high-ability peers; high-ability students may benefit because they are the ones engaged in giving the elaborated help. Of course, these interactions occur within the context of a school, class, or group environment where ambient mechanisms are still operating, so ambient mechanisms are never entirely without influence. We suspect that teachers seldom capitalise on heterogeneous grouping of students to encourage peer interaction and learning, and it may be that teachers need to foster an ambient environment or culture of help seeking and help giving in order to promote cooperative learning.

6.3 Accounting for Inconsistencies in Findings

It must be remembered that the compositional effects we have identified are small. In earlier chapters, we have gone to some lengths to point out that our observations of small effect sizes, particularly at the school and classroom levels, seem inconsistent with findings of descriptive studies (e.g., on school mix and streaming) that have not included measures of learning outcomes. Either the outcome data are right and compositional effects are quite small, or the outcome-based studies are not picking up the magnitude of the effects others have observed from more descriptive research – or perhaps some combination of these explanations is required. We now turn to explanations for these inconsistencies in findings.

One explanation is that the findings from descriptive studies are correct – there are peer effects – it is just that the effects have little consequences for learning over and above those that can be accounted for by individual characteristics of students. Recall that, by definition, a compositional effect requires that the aggregate characteristic of a student group (e.g., mean level of ability) significantly predicts a student's performance on outcome measures as well as or better than an array of individual characteristics of the student (usually including his or her ability). This is a stringent requirement. It may be that the nature of the students in a school, a class, or a group substantially influences the instructional and normative environment for learning but that the effects on learning outcomes are adequately accounted for by students' individual characteristics. In other words, the picture created by descriptive studies is simply a reflection of individual students in the composite.

Another explanation is that there are peer effects but that outcome-based studies of compositional effects underestimate them because the operative level of peer effects is at a smaller level of aggregation than is typically studied. It could be, for instance, as much theory suggests, that peer effects occur in small clusters of students. Data aggregated at the school or class level, or even at the group level, may not capture the relevant processes by which peer effects occur. This could be because they do not have the fidelity to pick up small effects or because there are counterbalancing forces – some 'groups' within the class or school are productive groups and some are not, so they cancel or reduce each others' effects. In socially well-mixed schools, for example, the effects of school mix would be cancelled out by student subcultures in which those of high prior achievement excelled, whereas those of lower prior achievement generated a culture of resistance and school failure. Gamoran (1992), drawing on Heyns (1986), provides an extended discussion of this issue and points out that the dual forces cannot easily be quantitatively disentangled. This hypothesis might be tested in multi-level studies that identified the performance of students from different social class or prior achievement backgrounds. However, if the only measure of mix is by measuring the mean of likely mix variables, then the compositional effect could be cancelled out.

Yet another explanation is that there are peer effects but that outcome-based studies of compositional effects underestimate them because they do not do justice to the reciprocal relationships between students, teachers, and school organisation and management. Instead, they partial out these 'peer effects' in the analyses of compositional effects, either statistically, as in correlational studies, or by design, as in experimental studies. For example, students in high-stream classes or in schools in high socio-economic areas might take more courses and more academic or demanding courses than students in low-stream classes or students in low socio-economic schools. Correlational studies may deliberately control for these differences by including curriculum differentiation as a factor in the regression, thus inadvertently partialling out variance in the outcome measure that is, in part, a 'peer effect'. Experimental studies may attempt to equate or match groups or classes so that the effects are not confounded with differences in instruction, or meta-analyses may discard studies that do not meet the requirements of group or class equivalence (see Lou et al., 1996; Slavin, 1987). In either case, these practices would minimise the effects of peers.

Similarly, there are likely to be overlaps between the effects of teacher practices and student composition (Bryk et al., 1990) and between effects of school policy and practice and student composition. Nash and Harker (1998) acknowledge the last problem this way:

> *Certain aspects of school policy and practice, for example, are associated with the composition of the student population.... It is difficult to attribute any of the resultant achievement difference to either 'student composition' or 'teaching practice' – they overlap in complex ways. For these reasons, some of the variance accounted for by the so-called contextual or compositional factor may, in fact, be an unknown combination of school effects and student effects (p. 52).*

Note that Nash and Harker retain a conceptual division between school effects and compositional effects but Thrupp (1999b) questions the utility of this distinction altogether. Thrupp is essentially arguing that, because the impact of school composition on classroom instruction and school organisation and management processes is seen to be so large, much of the 'school effect' will in fact be an effect of school composition.

Finally, it could be that there are peer effects but that the outcome-based studies fail to model them in a theoretically appropriate way. Studies of contextual and compositional influence on educational outcomes simply use an 'input-output' formulation and have, crucially, omitted to model how these influences act through the process of education. As we indicated in the previous chapter, assuming that a variable such as the social mix of a class has a directly interpretable influence on the outcomes of class members is of little value if the exact mechanism and necessary conditions are not specified in the model. For example, Coleman et al. (1966) raised the possibility that, while school mix could have a significant impact on school performance, it might also be the case that ethnic minority or working-class students suffer from low self-esteem in schools with predominantly middle-class students. Hence, the positive effects of being in predominantly middle-class schools could be cancelled out. Alternatively, the structure of schools might lead to counteracting effects at differing levels of school organisation. For instance, the effects of being placed in a low stream might be seen to counteract the beneficial effects of a high socio-economic school mix. This is because streaming may create compositional conditions at the class level that do not reflect those of the school as a whole. This concern over modelling holds for both conventional regression models and the more recent multi-level models. The size and nature of the influence of such effects is almost certain to vary, dependent on which of the many possible processes is operative.

6.4 Conclusion

A major implication of our model is that there can be major costs of attending to the higher levels of the model, believing that this has a high likelihood of translating into student learning gains. These higher layers must not be ignored, as there can be instances of impact, but in general the costs relate more to not dealing with the lower levels of the model. The critical trade-off for policy makers is between attending to school and classroom organisational practices versus improving what happens once the classroom door is closed. Whether a school attends to its mix, streams or not, reduces class sizes, implements composite or single-level classes, or has coeducational or single-sex classes appears less consequential than whether the school attends to the nature and quality of instruction in the classroom, whatever the within-class variability in achievement. The learning environments within the class, and the mechanisms and processes of learning that they foster, are by far the more powerful. Moreover, altering the composition of a class or group is one way of promoting peer effects, but it may not be enough. As will be shown in the next chapter, attention also needs to be given to more careful specification of curricula, higher quality teaching, and higher expectations that students can meet appropriate challenges – these occur once the classroom door is closed and not by reorganising which students are behind those doors.

The peer influences at the school level are more difficult to discover, primarily because they are a mix of direct and, more specifically, indirect effects. The indirect effects are more related to the level of homogeneity, usually on the basis of achievement, but in some cases on the basis of socio-economic status or, on the basis of race, which can serve as proxies for achievement. Too often, students' ethnicity or gender may serve as a status characteristic, particularly in situations where there is no sound basis for making judgements about another's ability. These characteristics then become 'proxies' for ability that determine a student's relative influence and learning in a group. Most studies of the effects of school mix say little or nothing about the possible range or nature of processes underlying the phenomenon (for an exception, see Thrupp, 1999b). It is likely that schools develop processes that reflect their socio-economic mix. Solidly middle-class schools can have strongly supportive student cultures that allow them to teach an academic, exam-based curriculum and to organise and manage themselves relatively smoothly. Working-class schools can be quite the opposite. Upper-class schools have, among their students, higher levels of prior attainment, more previous experience of school success, more regular school attendance, higher academic goals, higher socio-economic occupational aspirations and expectations, and less involvement in 'alienated' student subcultures. It is rare, however, to have schools so narrowly homogeneous that these prescriptions apply in a strong and direct manner. Our estimate of school-mix effects in New Zealand is between zero and eight percent, and probably just under four percent, of the total variance in student achievement. Most of the effects are indirect, being multiplied across the layers of school organisation to yield a small influence on student learning outcomes.

The model clearly has major implications for where to direct attention relating to policy in schools. If the major peer effects are within the classroom, then more attention to this level of analysis seems more likely to lead to enhanced learning outcomes than does more attention to the mix of peers at the school or classroom level.

6.5 Recommendations for Further Research

- There is need for research on compositional effects at each layer of school organisation and on how these effects are transformed at successive layers. We need to know the magnitude of the effects and we need empirically based descriptions of the processes underlying allocation of students to classes (or groups) as a function of school (or class) composition.
- Relatedly, questions of the effects of school- and class-level composition on student learning outcomes may not be appropriate issues for research; rather, research might better be targeted at the implications of school- or class-level composition for composition at lower levels of school organisation.
- There is need for further specification of the relative influences of family and school resources and peer effects on student learning outcomes. This issue is readily answerable at the school level, but it should also be addressed at the class and group level.
- There is need for research on the relationship between peer learning environments and the associated learning mechanisms and processes. We need to understand more clearly which environments promote which mechanisms and processes, and we need to know whether there are reciprocal relationships between these two layers of influence.

Chapter 7

MAXIMISING PEER EFFECTS

As we have stressed throughout our report, altering the composition of various groupings of students only changes (but hopefully increases) the probability that learning among peers will occur. It does not guarantee it. In some respects, then, compositional effects may be considered as 'covariates' of learning among peers and we do not want to suggest that successful learning will occur only when certain conditions are met. On the contrary, although a focus on lower layers of our conceptual model is associated with higher probabilities that learning will occur, we cannot claim that good teaching or good schools are solely a function of the composition of various groupings of students or even the size of these groupings. As Webb and Palinscar (1996) suggest in their input-process-outcome model of group processes in the classroom (see Chapter 2), there are a number of ways in which structuring groups and group work influence group processes and outcomes. Composition in terms of ability, socio-economic status, ethnicity, or gender is but one 'input characteristic'. Perhaps more important are other characteristics such as altering group rewards or incentives, preparing students for group work, structuring interaction in groups, having students reflect on their interaction processes, and of course, structuring the roles that teachers play.

In this chapter, we describe four instructional approaches that show how teachers can capitalise on peer effects to maximise learning. Our aim is to illustrate additional ways teachers can structure productive interaction between peers that leads to learning. All approaches draw from a theory of learning that emphasises the reflective and social nature of learning. All reflect a view of learning that is 'constructivist', where peers are seen as aiding in the construction of knowledge. All approaches have been chosen as representing sound and proven research programmes. Some aspects of the approaches, notably peer tutoring and reciprocal teaching, are in current use by teachers in New Zealand; others represent innovative changes to existing pedagogy.

The approaches emphasise the pivotal role of social processes, and the procedures to introduce and support these processes are based on and embody a coherent set of learning principles. The emphasis has moved from cognitive strategy instruction aimed at the individual student to building a classroom culture supportive of active knowledge construction. The aim is to turn over to students the high-level processes in learning by helping them to formulate their own goals, to do their own activating of prior knowledge, to ask their own questions, to direct their own inquiry, and to do their own comprehension monitoring. While recognising the need for teacher guidance and an authoritative knowledge role, the idea is to give children "a higher level of agency in the knowledge-building process" (Scardamalia & Bereiter, 1991, p. 40). In the examples presented below, this is accomplished by providing external, generally social, supports for higher-level cognitive processes, and then facilitating the transfer of these processes from expert to child control. Another common aim is to make meta-cognitive activity, which is normally covert, into something that can be publicly considered.

7.1 Reciprocal Teaching

Reciprocal teaching, developed by Palincsar and Brown (1984), is an instructional procedure to improve reading comprehension in students who are able to decode, but have difficulty in comprehending, age-level reading text. The teaching procedure focuses on the dialogic teaching (i.e. teaching through dialogue between teacher/leader and students) and guided practising of four reading comprehension fostering and monitoring strategies. These strategies are questioning, summarising, clarifying, and predicting. Working together with a small group of students reading a common text, the teacher thinks aloud and models the use of the four strategies so that the original covert comprehension fostering and monitoring processes are made visible to students. Initially, the teacher takes major responsibility for leading the discussion among the group. Then the students take turns to take over the teacher's role to lead the discussion for one segment of the text at a time. The teacher supports students' participation by prompting, praising, altering the demand on the students, or providing extra scaffolding when necessary so as to ensure students' successful participation.

The reciprocal teaching dialogue is structured by the use of the four strategies. The leader generates a question to which members in the group respond, summarises the segment of text, notes or solicits points to be clarified, and predicts about the content of upcoming text. The other group members interact with the leader by exchanging comments and elaborating upon the leader's contribution. According to Palincsar (1986b), the articulation of each of the four strategies has specific functions, but the integrative use of them can facilitate students' comprehension fostering and monitoring as a whole. Summarising helps students to identify and integrate the main points in the text. Questioning requires students to identify the kind of information that is important enough to provide the substance for a question. The question itself is used as a device for self-test to ascertain that students can indeed answer their own question. Clarifying helps students to be alert to any breakdowns in comprehension and to take the necessary actions to restore meaning. Finally, predicting helps students to link their prior knowledge with the new knowledge they encounter in the text as they guess what will be coming up next. It gives students a purpose for reading; that is, to confirm or disconfirm their prediction (Palincsar, 1986b).

Although the reciprocal teaching dialogue is the key component in all studies in the literature, two forms of reciprocal teaching have been adopted by Palincsar and Brown: (1) reciprocal-teaching-only (RTO); and (2) explicit-teaching-before-reciprocal-teaching (ET-RT). The two forms differ in how and when the initial instruction in the cognitive strategies takes place (Rosenshine & Meister, 1994). In the first form (RTO), all modelling and instruction in how to develop and apply the four cognitive strategies takes place during the dialogue (Palincsar & Brown, 1984). In the second form (ET-RT), the four strategies are taught separately during three to six traditional lessons that are conducted before the dialogue begins (Brown & Palincsar, 1989; Palincsar, Brown, & Martin, 1987; Palincsar, David, Winn, Stevens, & Brown, 1990). These extra lessons are included in order to "introduce the students to the 'language' of reciprocal teaching by providing direct instruction in each strategy" (Brown & Palincsar, 1989, p. 33). Rosenshine and Meister (1994) found no significant differences in effect size between the two forms of reciprocal teaching on either standardised tests or experimenter-developed tests of reading comprehension.

Research investigating reciprocal teaching has been conducted over the past fifteen years covering large numbers of teachers working primarily with remedial, special education, and at-risk students in first grade (Year 2) through secondary school in the United States (Brown & Palincsar, 1989; Palincsar, 1986a; Palincsar & Brown, 1984; Palincsar & Klenk, 1991), as well as in New Zealand (Fung, 1999; Gilroy & Moore, 1988; Kelly, Moore, & Tuck, 1994; Le Fevre, 1996; Westera & Moore, 1995). Rosenshine and Meister (1994) reviewed 16 quantitative studies on reciprocal teaching conducted in the years between 1984 and 1992, including the studies by Brown and Palincsar. They reported effect sizes of .32 when standardised tests were used to assess student comprehension performance and .88 when experimenter-developed comprehension tests were used, favouring reciprocal teaching.

The success of reciprocal teaching does not lie in the procedures that are part of the surface activities (namely, questioning, summarising, clarifying, and predicting), but in the quality of the interaction during these activities (see Brown & Campione, 1996). The essence of reciprocal teaching is its focus on thinking aloud the cognitive and meta-cognitive strategies involved in the reading process. As the teacher is thinking aloud, so are the students, creating a window on the way group members (be they expert or novice readers) are processing the text. The teacher's responsibility is to maximise the use of this window and to ensure the quality of the verbal interaction. On the one hand, the teacher constantly monitors and evaluates students' comprehension process and provides immediate feedback. On the other hand, the teacher consciously transfers the responsibility of the meaning-making process by creating opportunities for cognitive conflict between peers through encouraging students' verbalising or externalising their thoughts. During the activities of questioning and clarifying, students are engaged in quality interaction if timely, relevant, correct, and sufficiently elaborated help is given and received. During the activities of summarising and predicting, students are engaged in quality interaction if they continue to actively construct and reconstruct their understanding of the text based on the feedback they received.

Although reciprocal teaching is a teacher-directed group learning activity, its success also relies on peer interaction to facilitate the learning process. Palincsar and Brown (1988) observed that "peers are frequently in a better position to assist one another in comprehension activity, since they are more likely to be experiencing the same kind of difficulty in comprehending the text than [are] teachers, for whom comprehension occurs with relative automaticity" (p. 57). According to Damon (1984), peers speak at a level that students can easily understand, peers are more likely to challenge one another than to challenge the teacher, and communication between peers is less threatening than is an evaluative response from the teacher. Therefore, it is important for the teacher to gradually transfer the control of the discussion to students.

To avoid the possibility of creating a helper/helpee caste system that reduces participation by the less able, as observed by Ross and Cousins (1995) in group learning activities, it is important for the teacher to ensure that each student in the group has a chance to take over the role as the 'teacher', to lead the reciprocal dialogue. The use of role switching is more likely to promote desirable help seeking and help giving behaviours.

Taken together, when the quality of verbal interaction is ensured, reciprocal teaching provides a collaborative context between teacher and students, as well as among students, that facilitates the task of constructing meaning from text. With the assistance of the teacher and peers, students become increasingly proficient at applying the four comprehension strategies while reading. This procedure embodies a socio-cognitive perspective that contends that cognitive development occurs when concepts first learned through social interactions become internalised and are made one's own (Vygotsky, 1978).

7.2 Collaborative Reasoning

Collaborative reasoning is an instructional strategy designed to make classroom discussion more intellectually stimulating by capitalising on critical features of naturally occurring talk among peers. A major theme of this report is that talk among peers is a powerful influence on learning. Teachers recognise the importance of children's discussion, and structure numerous classroom opportunities – small group and whole class, formal and informal – to promote verbal engagement between peers. However, there is increasing concern that the patterns of interaction in some group discussions limit the opportunities for learning. Observational studies of classroom discussions have long noted the predictable I-R-E (initiate-respond-evaluate) pattern in which the teacher initiates a verbal interchange with a question, a student responds, and the teacher completes the interaction by offering an evaluation of the child's reply (Mehan, 1972; Sinclair & Coulthard, 1975). In such interactions, the teacher controls most aspects of the communication. It is the teacher who controls which children speak, how long they speak for, the sequence of speakers, and the elements of the discussion that will be furthered or discontinued. Students talk mostly with the teacher, rather than with each other. As a result, the combined word output of all children in a class discussion is typically less than that provided by the teacher alone (Cazden, 1988).

Collaborative reasoning, on the other hand, has a more complex pattern of interactions between students and the teacher. The discussion is created spontaneously by the group members, who may take the discussion in various unknown directions. Greater control of the discussion is in the hands of the students, allowing greater expressive latitude (Commeyras, 1994; Raphael & McMahon, 1994) intended to stimulate both personal engagement and critical thinking (Waggoner, Chinn, Yi, & Anderson, 1995). It has typically been used in the context of discussions about books or stories that children have been reading.

Collaborative reasoning is easily understood in the context of a discussion about a story that children have read. In a typical session, the teacher initiates the discussion with a single central question about a major debatable issue in the story. Children first make public their position on the issue, which could include being unsure, and then advance reasons and supporting evidence for their position. The evidence may be textual or it may be from prior knowledge and experience. The children listen carefully and evaluate the arguments of their peers. When there is disagreement, children are encouraged to challenge with counter-arguments. Through a process of weighing the reasons and evidence presented, children may strengthen or weaken their commitment to their original positions, or they may change their positions.

Collaborative reasoning discussions are characterised by open participation similar to a natural conversation. There is no need for children to raise hands to speak; participants speak one at a time, interruptions are kept to a minimum, and there is an even balance between the rights of the teacher and those of the children. During the discussion, the teacher provides scaffolding that promotes the development of independent thinking and student management of turn taking. The teacher may challenge a child's argument, ask for clarification, offer a counter-argument, or prompt for evidence to support a position – just like any other member of the group. In addition to their participation, teachers may also model reasoning processes by thinking out loud, acknowledging good reasoning in the children, providing summaries of what the children have said, and using the vocabulary of critical and reflective thinking. These teacher behaviours will depend on such things as the dynamics of a particular group, the direction of the discussion, or the degree of skill that the children have in their thinking strategies.

Recent interest in interactive argument within peer groups comes from at least three sources. First, as highlighted in Chapter 2, structuring controversy among students around curriculum materials invokes learning processes that heighten motivation and achievement in students (Johnson & Johnson, 1995). Second, shared reasoning in collaborative groups is a part of everyday life, whether it be children organising the events for a class picnic, or members of a school board debating the merits of a new building. Peer discussion featuring argument has been proposed as a way to develop reasoning not just in literature (Commeyras, 1994; Waggoner et al., 1995), but also in social studies (Onosko, 1990; Pontecorvo & Girardet, 1993), mathematics (Putnam, Lampert, & Peterson, 1990), and science (Cavalli-Sforza, Lesgold, & Weiner, 1992; Mason, 1998). Third, peer group interactive argument may be the primary means through which students learn to reason. Vygotsky (1978) has argued that reasoning processes become internalised by an individual through a process involving social interaction with others.

The nature of peer interaction is critical to the success of collaborative reasoning. The 'argument networks' (Chinn & Anderson, 1998) that emerge from peer discussions are constructed collaboratively, with many children contributing to the final web of ideas. Networks will be stronger to the extent that peers have negotiated and renegotiated meanings and ideas, through processes of critical opposition and co-construction, to share a new common knowledge that incorporates a variety of perspectives. Peer groups that are heterogeneous with regard to gender, ethnicity, and socio-economic status are more likely to elicit a broader range of perspectives that may be applied to an issue. This heterogeneity also increases the likelihood of modelling effects (see Chapter 2) from high-status peers. One of the goals of structured collaborative reasoning is for students to critically examine the perspectives of others. In operational terms, this might be seen in explicit counter-arguments developed during the discussion, or it might be seen in individual students being able to simultaneously hold multiple perspectives on an issue.

Research on collaborative reasoning has received most impetus from a series of studies conducted by Anderson and his colleagues at the University of Illinois (Anderson, Chinn, Chang, Waggoner, & Yi, 1997; Anderson, Chinn, Waggoner, & Nguyen, 1998; Chinn & Anderson, 1998; Nguyen-Jahiel, Anderson, Waggoner, & Rowell, 1998; Waggoner et al., 1995). Much of this research has focused on understanding the nature of peer group discourse that occurs during discussions about stories in reading lessons. For example, Anderson et al. (1997) analysed videotaped transcripts from 20 collaborative reasoning discussions in fourth-grade (Year 5) classrooms to examine the effectiveness of children's naturally occurring arguments about stories the children had read. Although the logic in most children's discussions was flawed by conventional standards of analysis, the arguments were coherent and effective when judged against a common knowledge base about the issue and an ongoing discussion in which participants were active and cooperative. Thus, children as young as nine years were able to participate in collaborative reasoning in classrooms. Subsequently, Chinn and Anderson (1998) presented two methods for representing the macrostructure of these discussions – argument networks and causal networks – and suggestions were made for how teachers could promote greater development of reasoning skills through use of these networks.

Perhaps the greatest insight into the nature of the discourse about stories comes from a comparison of discussions occurring in two types of discussions (Anderson et al., 1998). High- and low-ability reading groups in four classrooms participated in conventional 'recitation' and collaborative reasoning discussions about stories. Collaborative reasoning resulted in greater student engagement and interest during story discussions. For example, the rate of student talk under the collaborative reasoning structure was double that under the recitation structure, while teacher talk declined by approximately one-third. Consecutive student turns were more than seven times more likely in collaborative reasoning discussions than in conventional discussions, while the frequency of questions from the teacher was almost halved. The nature of teacher questions also changed, with teachers being only one-sixth as likely to ask 'information-checking' questions, yet almost twice as likely to ask 'information-seeking' questions during collaborative reasoning discussions. Student interjections and back-channelling comments about other speakers' views both increased as group members competed for entry into the more conversational style of the collaborative discussions. These findings support the conclusion that student engagement is enhanced in collaborative reasoning.

The relationship between the discourse and the story text also differed under the two instructional conditions. Students were six times more likely to make explicit references to textual information during a collaborative reasoning discussion, usually to strengthen their arguments. On the other hand, as a function of having less need to prompt students to remember story information or to ask questions that constrain student answers to the story, teachers were almost six times less likely to make explicit references to the text in collaborative discussions. When teachers did use text information, it was to promote, reward, or model good reasoning strategies. Few utterances could be construed as arguments in conventional discussions, whereas during collaborative reasoning most utterances expressed arguments, challenged the arguments of their peers, or responded to challenges with elaboration and supporting information. Finally, children in collaborative reasoning were much more likely to express a personal affective response to an aspect of the story, and for this response to be directed towards their peers rather than towards the teacher. As a consequence, children were more likely to look at each other when they spoke and to use each other's names. These findings reflect a pattern of 'knee-to-knee, eye-to-eye' interaction characteristic of high engagement.

In summary, collaborative reasoning is an instructional method that changes the nature of student engagement in the peer discussions that influence student learning. It draws students into collaboration with their peers in the construction of knowledge by capitalising on peer influences as they exist in naturally occurring discussions. Designed to promote greater reasoning skill, the method also leads to greater and more positive engagement in learning. By focusing on the type of engagement seen in natural conversations, collaborative reasoning represents a tutorially configured environment that mimics an ambient environment. This highlights the difficulty of examining the effects and the learning processes and mechanisms of one environment separately from the other. Although still in its infancy, the method offers much promise for a type of discourse that operates across all curriculum areas.

7.3 Fostering a Community of Learners

The Fostering a Community of Learners (FCL) programme is an interactive classroom learning environment (Brown & Campione, 1990, 1996) combining several components. It relies on peer interaction to achieve its learning goals. The interactive components or practices contribute synergistically to the achievement of the environment (Brown & Campione, 1996) and are underpinned by a learning theory that draws on both cognitive and socio-cultural research traditions.

A main goal of the FCL programme is to create a learning community that features collaboration and 'distributed expertise' (Brown et al., 1993; Brown & Campione, 1994). In such a community, knowledge is seen as situated within the activities engaged in by a community. Rather than residing statically within an individual's mind, cognition is borne of participation in culturally organised practices with other people, in using the tools (e.g., books, computers, science equipment, videos) of the community. Thus, participation in social practices is seen as the key to the generation of learning and new understanding. Although all students are responsible for sharing a common body of knowledge and practice, expertise often becomes distributed as individuals specialise in a particular aspect of interest related to the common theme. The whole classroom community ultimately benefits from the varieties of expertise that develop as a function of engagement in joint activities (Brown & Campione, 1996).

The activities involved in the FCL programme emphasise research practices, such as discourse, argumentation, reading, writing, and computer use, in the pursuit of shared understanding of in-depth disciplinary (science) content (Brown et al., 1993; Brown & Campione, 1994). An FCL unit is organised around a main theme and consists of a research-share-perform cycle that lasts for approximately eight to 10 weeks. The research practices are supported by the use of familiar, recursive participant structures (Brown et al., 1993; Brown & Campione, 1994), through which students engage collaboratively with one another. The core collaborative participant structures of the FCL programme are: reciprocal teaching (Brown et al., 1993; Brown & Campione, 1990; Palincsar & Brown, 1984), jigsaw teaching activities (Brown et al., 1993; Brown & Campione, 1994), crosstalk (Brown & Campione, 1994), and cross-age tutoring (Brown & Campione, 1994, 1996). All four participant structures require that students share their developing knowledge and expertise with classmates. In addition, each research unit is set up so that sharing and collaboration are not just pleasant aspects of the programme but are essential to achieving the learning goals and to performing the final assessment activities for each unit.

Reciprocal teaching

This is designed to assist students in monitoring their comprehension during reading as discussed above (Palincsar & Brown, 1984). Reciprocal teaching serves as an integral part of the research cycle engaged in by the class and a principal form of collaborative learning in an FCL classroom (Brown et al., 1993; Brown & Campione, 1996). This procedure fosters explicit consideration of basic comprehension monitoring strategies (questioning, summarising, clarifying, and predicting), and when used in a research group situation, supports student discussion and understanding of research material (Brown et al., 1993). A typical reciprocal teaching session in an FCL classroom involves six or so participants, often including a teacher, parent, or older student, each of whom takes a turn at leading a discussion about an article or other research material they need to digest as a group (Brown et al., 1993; Brown & Campione, 1994, 1996). After a section has been read, the leader begins the discussion by asking a question about what has been read. Once members have had their say, the leader summarises, identifying the gist of the discussion. As comprehension problems arise, the members seek clarification from the group. From time to time, the leader will ask for predictions about future content. The goal is to negotiate the meaning of the text so that all members reach a consensus in understanding. This procedure is designed so that readers of differing abilities can participate at levels appropriate to their expertise. Students not yet capable of full participation benefit from the modelling of the comprehension monitoring strategies by more expert participants (Brown & Campione, 1994, 1996). Reciprocal teaching sessions are used as necessary throughout the research cycle.

Jigsaw teaching activities

As part of the research process engaged in within the FCL classroom, groups of students form around a common goal of exploring a subtopic of interest that emerges from the main whole class theme. Typically, five or six subtopics develop, resulting in five or six research groups. The various research groups that form become responsible for mastering the subtopic, and subsequently for teaching it to members of the other research groups. This is organised with the use of an adapted jigsaw method of cooperative learning (Brown et al., 1993; Brown & Campione, 1994, 1996), whereby students from each research group re-form into jigsaw learning groups. In these learning groups, each student is the holder of knowledge pertaining to his or her subtopic and must teach it, using reciprocal teaching methods, to the other members so that each member of the jigsaw group comes away with knowledge of each piece (subtopic) in the overall research puzzle (theme).

Crosstalk

Unlike reciprocal teaching and jigsaw, crosstalk is a whole-class activity. During the research cycle, students require opportunities to come together as a class to discuss where they are at in terms of their developing understanding before they are required to teach it to others. Members of a particular research group report to the class their findings to date, at which time students from other research groups ask questions of clarification or request additional information. Whole class discussion ensues where groups talk across groups, hence the term 'crosstalk'. These opportunities make transparent to research groups where gaps in their understanding lie. Faced with questions they cannot answer, students take notes of the knowledge or explanations lacking and return to their research groups for a new cycle of research. Crosstalk essentially serves as a group comprehension monitoring device, and prepares individual research group members for jigsaw where they must separate from their fellow members and teach their material to the jigsaw group members in the absence of their research group.

Cross-age tutoring

Further opportunities to talk about learning are provided through cross-age tutoring sessions, conducted both face-to-face and via electronic mail. The FCL programme has involved fifth to seventh grade (Year 6–8) students working with younger (first to third grade/Year 2–4) students (Brown & Campione, 1994, 1996). The older students, who are experienced in reciprocal teaching processes, know the content area of interest and are trained in tutoring the younger students. Tutors work with two tutees, helping them in all aspects of the research process as needed. For example, they facilitate reading, discussing important information, writing and editing, and establishing new learning goals. The older students also act as mentors and monitors during reciprocal teaching sessions and jigsaw activities, providing leadership expertise and role models of practice. Brown and Campione (Brown & Campione, 1996) maintain that the provision of older students to assist the research efforts of younger students not only fosters self-esteem among the older ones, but provides individualised instruction to younger students. It also relieves the teaching burden on the classroom teacher and reinforces the collaborative structures within the community as a whole.

The success of the FCL programme has been measured using a mixture of qualitative and quantitative methodologies (Brown, 1992; Brown & Campione, 1994, 1996). No attempt is made to isolate the effects of any one aspect of the programme, which is seen as functioning as a systemic whole. Outcome measures include traditional pre-test and post-test data on both experimental and control group students, combined with micro-genetic analyses of several children or a group.

Fifth and sixth grade (Year 6–7) students in the FCL research classroom were compared with a partial control group. Both groups received the same treatment for the first semester, but the partial control returned to a traditional science classroom setting for the remainder of the school year (Brown & Campione, 1994). The partial control students had access to the same books, videos, computers, and so on as the research classroom group. These two groups were compared to a full control group who read the key materials for each unit but did not engage in research practices. Short answer quizzes for each of three science units were given before and after each unit was introduced. The results showed that, although the FCL research group and the partial control group did not differ from each other in science knowledge on unit 1 (semester 1), they outperformed those in the full control group. The students in the research group showed substantial increases in knowledge compared with those in both the full control and partial control groups after completing both units 2 and 3.

Similar benefits of participating in the collaborative research classroom were reported in terms of dynamic assessments of critical thinking about content (Brown, 1992). On tasks assessing 'novel application' of science concepts, students who had participated in an FCL research classroom were reported to be more adept than control group students at responding both with novel variations of science principles taught and with truly innovative ideas. Clinical interviews and application, and transfer tests assessing critical and flexible use of information learned, showed that research classroom participants improved considerably in these regards over the course of their involvement in the FCL classroom (Brown & Campione, 1994).

As the FCL programme is both a science curriculum and an integrated literacies programme, measures of reading comprehension have also been obtained (Brown & Campione, 1994). To measure the effects of participation in the programme on reading comprehension in general and to examine whether skills developed transfer to other contexts, data were obtained on students' reading comprehension on materials unrelated to the science domain of study. Comparisons were made between students who participated in an FCL classroom for a full year, those who participated in an FCL classroom for one semester (partial control group), those who did not participate in an FCL classroom but who engaged in reciprocal teaching activities, and those who read the same science curriculum materials as the students in the FCL class. Criterion-referenced tests of fifth and sixth grade (Year 6–7) students' reading comprehension, given at the beginning and end of the year, showed that students who spent a full year in a FCL research class made higher gains in reading comprehension than did those in the reciprocal-teaching-only group. This was despite the fact that the reciprocal-teaching-only group received more practice in the procedures and in taking tests of that nature. Both the read-only control group and the partial control group failed to show the same gains obtained by the FCL research group. Furthermore, analyses of a range of reading comprehension question types (fact, inferential, gist, and analogy) showed that FCL research group students improved substantially in their ability to answer appropriately all but the fact-based questions on which they scored well at the start.

With regard to argumentation skills, analyses of FCL student dialogues have shown that the use of advanced analogical reasoning and causal explanations increased over time, contributing to the development of more complex argumentation formats (Brown & Campione, 1994). FCL students showed a more sophisticated use of evidence and explanatory devices, such as negative evidence and warrants, in supporting their arguments.

Brown and Campione (Brown & Campione, 1994, 1996) argue that the success of their programme lies in adherence to a set of first principles of learning, central to which is the establishment of a collaborative community of learners. Through the practices and activities described, a dialogic base is established where student-to-student as well as student-to-adult discourse is fostered (Brown & Campione, 1994). Within this dialogic context, knowledge is shared and expertise becomes distributed, engendering a community of research practice.

7.4 Computer Supported Intentional Learning Environments (CSILE)

Computer supported intentional learning environments (CSILE), as the name suggests, is an effort to support intentional learning by students using computer technology. Intentional learning, as conceived by the principal investigators, Marlene Scardamalia and Carl Bereiter, involves students investing effort in learning over and above what they invest in immediate school tasks (Bereiter & Scardamalia, 1989). Intentional learning involves having goals, rather than strategies, and having knowledge as a goal. Often the strategies that students have, although they help them accomplish school tasks, actually work against the very intellectual activities school tasks were intended for. The authors see learning – in cognitive terms the construction of knowledge by students – as a by-product of school work, rather than something students actively attend to (Scardamalia, Bereiter, & Lamon, 1994).

The aim of the programme is to bring knowledge to the fore by having a medium "in which knowledge could be objectified, represented in an overt form so that it could be evaluated, examined for gaps and inadequacies, added to, revised and reformulated" (Scardamalia et al., 1994. p. 201). CSILE is basically a hyper media system built around a student-generated database. The work students do in various subjects is entered as text or graphical notes into a common database where everything students generate is available to others, so that commenting on knowledge building becomes a normal central activity. The system aimed to make the construction of knowledge a social activity. The student-generated database objectifies the accumulating knowledge of the classroom group. The features of the technology support joint planning, sharing of findings, and commenting and making suggestions to other students. Increasingly, the project led the instigators to see intentionality as something that can exist at a group level and to ask whether classes can have a goal of understanding and whether achieving an understanding is more than simply a tabulation of what individual students learn.

An aim of the research was to design a learning environment that transferred the scientific model of knowledge building, namely the construction of a public understanding of things, to the classroom. In advocating an educational focus on the social construction of knowledge, the authors implied that this has advantages for the individual learner. The claim is that cognitive processes that are used to articulate beliefs and ideas in social interaction are then available to the individual student for self-reflection. In particular, the researchers were concerned with the development of higher-level learning goals such as question asking, hypothesising, and explaining.

With respect to question asking, they asked whether children can be appropriately scaffolded to ask not simply text-based questions, but more educationally valuable, knowledge building questions. This is on the assumption that deeper processing is required to produce a situation model that represents the world as it relates to text than is required for a text-based model that is a semantic representation of what the text says. Further, the authors identified that there was substantial educational wisdom available in a group with respect to useful questions to ask to build knowledge, even if they knew very little about the topic in hand. The learning environment set up aimed to capitalise on this by instigating a social process that allowed the best judgements to come to the fore.

There is empirical evidence in the work of Scardamalia and Bereiter for claims of higher-level cognitive processing as a result of working in the computer supported intentional learning environment. The assessments that have been applied in evaluating CSILE range from standardised tests to purpose-built instruments. The use of CSILE originally focused on language domains, and the results on standardised tests (Canadian Test of Basic Skills) reveal a significant effect on language measures of experience on CSILE. Such results, while indicating some general effects favouring classrooms where CSILE was used, do not provide any evidence of causal factors.

Some of the results from evaluation measures designed to tap those higher-level thinking processes CSILE is meant to foster, are more indicative of the mechanisms that might support enhanced thinking and learning. Data from ratings of written pieces suggest that students in the CSILE classes scored better on indices relevant to knowledge building, in particular, depth of explanation. Similarly, using diagrams and conceptual measures of explanation, it was found that CSILE students produced significantly more advanced explanations and diagrams that contained more causal/dynamic information. A comparison of student comments on their own and others' portfolio selections suggests that CSILE students, to a significantly greater extent, went progressively deeper into describing what their work was about. They more often cited learning goals in their reasons for selecting a particular piece of work, and gave detailed descriptions of what they had learned from doing the work. They were more reflective about their own knowledge and learning than non-CSILE students and also made more reflective comments about others' work.

Further evidence of the likely causal processes comes from a naturally occurring contrast between the ways two highly competent teachers used the CSILE environment (Bereiter & Scardamalia, ND). Because of confounding likely from teacher differences, causal relations are still only speculative. The models of use were described as an independent research model and a collaborative knowledge building model. The resulting data tracked from the computer use illustrated the differences in the relationship between process variables and learning outcomes. Significant differences were found between classrooms in four categories of variable, namely, productivity, exploring work of others, collaborating, and advanced knowledge processes. There was an overall higher level of CSILE use in the collaborative classroom with the exception that notes produced by the independent research model class were much longer. There were significant differences between the groups in constructive effort and knowledge quality and in the tendency to ask questions of explanation. Predictors of knowledge quality in both classes differed. In the independent research class, they were productivity and advanced processes (use of thinking type icons) and, in the collaborative class, the significant predictors were exploring and collaborating, two measures of communal activity.

As a designed learning environment, the computer supported knowledge medium appeared to support intentional learning and the creation of cooperative knowledge building. The resulting communal activity was significantly related to enhanced learning outcomes.

7.5 Conclusion

These four learning innovations, Reciprocal Teaching, Collaborative Reasoning, Fostering a Community of Learners, and Computer Supported Intentional Learning Environments, derive from socio-cognitive and socio-cultural theories of learning. They illustrate different ways researchers have manipulated the learning environments to capitalise on peer effects. The researchers in these innovative programmes stress that the nature of peer interaction is critical. They regard knowledge as situated in the activities engaged in by the community of learners. There is evidence from these examples that these innovations promote collaborative processes, broadly conceived, and that these processes are linked to proximal indicators of learning such as increased interest and student engagement, increased production of higher-level cognitive processes, and enhanced learning outcomes. Policy directed at this level of analysis seems more likely to lead to enhanced learning outcomes than policy directed solely at the composition of students in schools and classrooms.

REFERENCES

Abelson, R. P. (1995). *Statistics as principled argument.* New Jersey: Lawrence.

Aboud, F. E., & Mendelson, M. J. (1996). Determinants of friendship selection and quality: Developmental perspectives. In W. M. Bukowski & F. N. A, et al. (Eds.), *The company they keep: Friendship in childhood and adolescence* (pp. 87-112). New York, NY: Cambridge University Press.

Achilles, C. M., Kiser-Kling, K., Owen, J., & Aust, A. (1994). *Success starts small: Life in a small class. Final report.* Greensboro, NC: University of North Carolina.

Alexander, K. L., Cook, M. A., & McDill, E. L. (1978). Curriculum tracking and educational stratification: Some further evidence. *American Sociological Review, 43*, 47-66.

Alexander, K. L., McDill, E. L., Fennessey, J., & D'Amico, R. J. (1979). School SES influences: Composition or context? *Sociology of Education, 52*, 222-237.

Alexander, K. L., & Pallas, A. M. (1983). Private schools and public policy: New evidence on cognitive achievement in public and private schools. *Sociology of Education, 56*, 170-182.

Allington, R. L. (1977). If they don't read much, how they ever gonna get good? *Journal of Reading, 21*, 57-61.

Allington, R. L. (1980a). Teacher interruption behaviors during primary-grade oral reading. *Journal of Educational Psychology, 72*, 371-374.

Allington, R. L. (1980b). Poor readers don't get to read much in reading groups. *Language Arts, 37*, 872-876.

Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal, 83*, 548-559.

Allington, R. L. (1984). Content coverage and contextual reading in reading in groups. *Journal of Reading Behavior, 16*, 85-96.

Alpert, J. L. (1974). Do teachers adapt methods and materials to ability groups in reading? *California Journal of Educational Research, 26*, 120-123.

Alpert, J. L. (1975). Do teachers adapt methods and materials to ability groups in reading? *California Journal of Educational Research, 26*, 120-123.

Alton-Lee, A. G. (1984). *Understanding learning and teaching: An investigation of pupil experience of content in relation to immediate and long term learning.* Unpublished doctoral dissertation, University of Canterbury, Christchurch, NZ.

Alton-Lee, A. G., & Nuthall, G. A. (1990). Pupil experiences and pupil learning in the elementary classroom: An illustration of generative methodology. *Teaching and Teacher Education, 6*(1), 27-45.

Alton-Lee, A. G., Nuthall, G. A., & Patrick, J. (1987). Take your brown hand off my book: Racism in the classroom. *SET, 1*, item 8. Wellington: New Zealand Council for Educational Research.

Alton-Lee, A. G., Nuthall, G. A., & Patrick, J. (1993). Reframing classroom research: A lesson from the private world of children. *Harvard Educational Review, 63*(1), 50-84.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79*, 193-273.

Anderson, R. C., Chinn, C., Chang, J., Waggoner, M., & Yi, H. (1997). On the logical integrity of children's arguments. *Cognition and Instruction, 15*(2), 135-167.

Anderson, R. C., Chinn, C., Waggoner, M., & Nguyen, K. (1998). Intellectually stimulating story discussions. In J. Osborn, F. Lehr, & et al. (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 170-186). New York, NY: The Guilford Press.

Anderson, R. C., Wilkinson, I. A. G., & Mason, J. M. (1991). A microanalysis of the small-group, guided reading lesson: Effects of an emphasis on global story meaning. *Reading Research Quarterly, 26*, 417-441.

Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry, 11*, 3-42.

Arai, A. B. (1999). Homeschooling and redefinition of citizenship. *Education Policy Analysis Archives, 7*(27), (http://epaa.asu.edu/epaa/v7n27.html).

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management, 15*(4), 623-645.

Baker, D. P., & Stevenson, D. L. (1986). Mothers' strategies for children's school achievement: Managing the transition to high school. *Sociology of Education, 59*(1), 156-166.

Ball, S. J. (1981). *Beachside comprehensive*. Cambridge: Cambridge University Press.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs: NJ: Prentice Hall.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*, 1175-1184.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality & Social Psychology, 71*(2), 364-374.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*, 593-604.

Barr, R. (1982). Classroom reading instruction from a sociological perspective. *Journal of Reading Behavior, 14*(4), 375-399.

Barr, R. (1988). Conditions influencing content taught in nine fourth-grade mathematics classrooms. Elementary School Journal. *Elementary School Journal, 88*(4), 387-411.

Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago: University of Chicago Press.

Barr, R., & Dreeben, R. (1991). Grouping students for reading instruction. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 885-910). New York: Longman.

Bates, P., et al. (1992). *Beyond tracking*. Ann Arbor, MI: Progress for Educational Opportunity, Michigan University.

Beauregard, M., & Ouellet, G. (1995). Development and testing of a dropout prevention program centered around extracurricular activities. *Loisir & Societe & Leisure, 18*(2), 373-394.

Bellanca, J., & Swartz, E. (Eds.). (1993). *The challenge of detracking: a collection*. Palatine, IL: IRI/Skylight Publishing (ERIC Document Reproduction Service No. ED 369 573).

Bentz, J. L., & Fuchs, L. S. (1996). Improving peers helping behavior to students with learning disabilities during mathematics peer tutoring. *Learning Disability Quarterly, 19*(4), 202-215.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick, et al. (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum Association.

Bereiter, C., & Scardamalia, M. (ND). *Two models of classroom learning using a communal database* (Technical Report): Ontario Institute for Studies in Education.

Berger, J., Cohen, B. P., & Zelditch, J. M. (1972). Status conceptions and social interaction. *American Sociological Review, 37*(3), 241-255.

Berger, J., Rosenholtz, S. J., & Zelditch, M. (1980). Status organizing processes. *Annual Review of Sociology, 6*, 479-508.

Berliner, D. C. (1981). Academic learning time and reading achievement. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* . Newark, DE: International Reading Association.

Berndt, T. J. (1979). Developmental changes in conforming to parents and peers. *Developmental Psychology, 15*, 608-616.

Berndt, T. J., Hawkins, J. A., & Jiao, Z. (1999). Influences of friends and friendships on adjustment to junior high school. *Merrill-Palmer Quarterly, 45*(1), 13-41.

Berndt, T. J., & Keefe, K. (1995). Friends' influence on adolescents' adjustment to school. *Child Development, 66*, 1312-1329.

Bershon, B. L. (1992). Cooperative problem solving: A link to inner speech. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 36-48). New York: Cambridge University Press.

Betts, J. R., & Shkolnik, J. L. (1999). The behavioral effects of variations in class size: The case of math teachers. *Educational Evaluation & Policy Analysis, 21*(2), 193-213.

Bevington, J., & Wishart, J. G. (1999). The influence of classroom peers on cognitive performance in children with behavioral problems. *British Journal of Educational Psychology, 69*(Part 1), 19-32.

Bidwell, C., & Kasarda, J. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education, 88*, 401-430.

Bigelow, B. (1993). Getting off track: Classroom examples for an anti-tracking pedagogy. *Rethinking Schools, 7*(4), 18-20.

Bishop, J. H. (1989). Why the apathy in American high schools? *Educational Researcher, 18*, 6-10.

Bloome, D., & Green, J. (1984). Directions in the sociolinguistic study of reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 395-421). New York: Longman.

Blumenfeld, P., Pintrich, P. R., Meece, J., & Wessels, K. (1982). The formation and role of self-perceptions of ability in elementary school classrooms. *Elementary School Journal, 82*, 401-420.

Blumenfeld, P. C., Hamilton, V. L., Bossert, S. T., Wessels, K., & Meece, J. (1983). Teacher talk and student thought: Socialization into the student role. In J. M. Levine & M. C. Wang (Eds.), *Teacher and student perceptions: Implications for learning* (pp. 143-192). Hillsdale, NJ: Erlbaum.

Bode, R. K. (1993). *Hierarchical linear modeling of class ability range on student mathematics achievement.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA (ED 360 317).

Boivin, M., & Hymel, S. (1997). Peer experiences and social self-perceptions: A sequential model. *Developmental Psychology, 33*(1), 135-145.

Bondi, L. (1991). Attainment in primary schools. *British Educational Research Journal, 17*(3), 203-217.

Bondy, E. (1990). Seeing it their way: What children's definitions of reading tell us about improving teacher. *Education. Journal of Teacher Education, 41*, 33-45.

Boozer, M., & Rouse, C. (1995). *Intraschool variation in class size: Patterns and implications.*

Borg, G. (1999). *Group assessment processes and group gender composition.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Bossert, S. T. (1988-89). Cooperative activities in the classroom. In E. Z. Rothkopf (Ed.), *Review of research in education* (Vol. 15, pp. 225-250). Washington, DC: American Educational Research Association.

Bossert, S. T., Barnett, B. G., & Filby, N. N. (1984). Grouping and instructional organization. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 39-51). New York: Academic Press.

Bourdieu, P., & Paseron, J. C. (1977). *Reproduction in education, society and culture.* London: Sage.

Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices and student achievement. *American Educational Research Journal, 23*, 558-571.

Bowermaster, M. (1978). Peer tutoring. *The Clearing House, 52*, 59-60.

Braddock, J. H. (1990). *Tracking: Implications for student race-ethnic subgroups* (Technical Report No. 1). Baltimore: Center for Research on Effective Schooling for Disadvantaged Students.

Brewer, D. J., Krop, C., Gill, B. P., & Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Evaluation & Policy Analysis, 21*(2), 179-192.

Brewer, D. J., Rees, D. I., & Argys, L. M. (1995). Detracking America's Schools: The reform without cost? *Phi Delta Kappan, 77*(3), 210-215.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51*, 5-32.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188-228). New York: Cambridge University Press.

Brown, A. L., & Campione, J. C. (1990). Interactive learning environments and the teaching of science and mathematics. In M. Gardner, J. G. Greeno, R. Reif, A. H. Schoenfeld, A. DiDedda, & E. Stage (Eds.), *Toward a scientific practice of science education* (pp. 111-139). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229-270). Cambridge, MA: Bradford Book, MIT Press.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Mahwah, NJ: Lawrence Erlbaum Associates.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 393-451). Hillsdale, NJ: Lawrence Erlbaum.

Brown, B. B., Clasen, D. R., & Eicher, S. A. (1986). Perceptions of peer pressure, peer conformity dispositions, and self-reported behavior among adolescents. *Developmental Psychology*, 22, 521-530.

Brown, P., & Goren, P. (1993). *Ability grouping and tracking: Current issues and concerns. Achieving national goals*: ED 406 458.

Bruffee, K. A. (1973). Collaborative learning: Some practical methods. *College English, 34*(5), 634-643.

Bryk, A., Lee, V., & Smith, J. (1990). High school organisation and its effects on teachers and students: An interpretative summary of the research. In W. Clune & J. Whitte (Eds.), *Choice and control in American education* (Vol. I,). New York: Falmer Press.

Bryk, A. S., & Driscoll, M. E. (1988). *The high school as community: Contextual influences and consequences for students and teachers.* Madison, Wisconsin: National Center on Effective Secondary Schools, University of Wisconsin.

Bryk, A. S., & Lee, V. E. (1992). Is politics the problem and markets the answer? An essay review of politics, markets, and America's schools. *Economics of Education Review, 11*(4), 439-451.

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good.* Cambridge, Massachusetts: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* London: Sage Publications.

Bukowski, W. M., Hoza, B., & Boivin, M. (1993). Popularity, friendship, and emotional adjustment during early adolescence. In e. a. B. Laurse (Ed.), *Close friendships in adolescence. New directions for child development* (Vol. 60, pp. 23-37). San Francisco, CA: Jossey-Base.

Bukowski, W. M., Newcomb, A. F., & Hartup, W. W. (Eds.). (1998). *The company they keep: Friendship in childhood and adolescence.* New York, NY: Cambridge University Press.

Burns, R. B., & Mason, D. A. (1995). Organizational constraints on the formation of elementary school classes. *American Journal of Education, 103*(2), 185-212.

Caldas, S. J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *Journal of Educational Research, 90*(5), 269-277.

Calvery, R., et al. (1992). *The difference in achievement between home schooled and public schooled students for grades four, seven, and ten in Arkansas.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN.

Camarena, M. (1990). Following the right track: A comparison of tracking practices in public and Catholic schools. In R. Page & L. Valli (Eds.), *Curriculum differentiation: Interpretive studies in U. S. secondary schools* . Albany, NY: State University of New York Press.

Camp, B. W., & Zimet, S. G. (1975). Classroom behavior during reading instruction. *Exceptional Children, 42*(2), 109-110.

Camp, W. (1990). Participation in student activities and achievement: A covariance structural analysis. *Journal of Educational Research, 83*(5), 272-278.

Campbell, A. L. (1965). A comparison of the effectiveness of two methods of class organization for the teaching of arithmetic in junior high school. *Dissertation Abstracts International, 26*, 813-814 (University Microfilms No. 65-6726).

Carnegie Foundation for the Advancement of Teaching. (1992). *School Choice: A special report.* Princeton, NJ: Author.

Carroll, A., Durkin, K., Hattie, J., & Houghton, S. (1997). Goal setting among adolescents: A comparison of delinquent, at risk, and not at-risk youth. *Journal of Educational Psychology, 89*, 441-450.

Carver, R. P. (1993). The case against statistical significance, revisited. *Journal of Experimental Education, 61*(4), 287-294.

Cavalli-Sforza, V., Lesgold, A. M., & Weiner, A. W. (1992). *Strategies for contributing to collaborative arguments: Procedings of the fourteenth annual conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Cazden, C. B. (1988). *Classroom discourse: The language of teaching and learning.* Portsmouth, NH: Heinemann.

Chalmer, J. B., & Townsend, M. A. R. (1990). The effects of training in social perspective taking on socially maladjusted girls. *Child Development, 61*(1), 178-190.

Chamberlain, G. S. (1993). Teachers of reading. In H. Wagemaker (Ed.), *Achievement in reading literacy* . Wellington, NZ: Research Section, Ministry of Education.

Chamberlain, G. S. (1997). The teachers. In R. A. Garden (Ed.), *Mathematics and science performance in middle primary school: Results from New Zealand's participation in the third international mathematics and science study* (pp. 177-212). Wellington, NZ: Research and International Section, Ministry of Education.

Chase, C. I., Mueller, D. J., & Walden, J. D. (1986, December). *PRIME TIME: Its impact on instruction and achievement. Final report.* Indianapolis, IN: Indiana Department of Education.

Chase, P., & Doan, J. (Eds.). (1994). *Full circle: New look at multi-age education.* Portsmouth, NH: Heineman Publishers. ED 371 864.

Chasin, G., & Levin, H. M. (1995). Thomas Edison accelerated elementary school. In J. Oakes & K. Hunter Quartz (Eds.), *Creating new educational communities, schools, and classrooms where all children can be smart* (pp. 130-146). Chicago, IL: University of Chicago Press.

Chen, X. (1997). *Students' peer groups in high school: The pattern and relationships to educational outcomes. Statistics in brief.* Washington, DC: National Center for Education Statistics, Department of Education.

Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions intended to promote reasoning. *Teachers College Record, 100*, 315-368.

Chinn, C. A., Waggoner, M. A., Anderson, R. C., Schommer, M., & Wilkinson, I. A. (1993). Situated action in the small-group reading lesson: A microanalysis of oral reading error episodes. *American Educational Research Journal, 30*, 361-392.

Chubb, J., & Moe, T. (1990). *Politics, markets, and America's schools.* Washington D.C.: The Brookings Institute.

Cignetti, M. J. (1974). A study of intraclass grouping and traditional groupings on students' terminal achievements during the last nine weeks in first semester type-writing (Doctoral dissertation, University of Pittsburgh). *Dissertation Abstracts International, 35*, 5765A.

Cocklin, B., & Battersby, D. (1987). Gender, the hidden agenda: A case study in educational decision-making. *New Zealand Journal of Educational Studies, 22*(1), 59-70.

Cohen, D. K., & Ball, D. L. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12*(3), 331-338.

Cohen, E. (1972). Sociology and the classroom. *Review of Educational Research, 42*, 441-452.

Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1-35.

Cohen, E. G. (1973). Modifying the effects of social structure. *American Behavioral Scientist, 16*, 861-879.

Cohen, E. G. (1982a). Expectation states and interracial interaction in school settings. *Annual Review of Sociology, 8*, 209-235.

Cohen, E. G. (1984). Talking and working together: Status interaction and learning. In P. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *Instructional groups in the classroom: Organization and processes* (pp. 171-188). Orlando: Academic.

Cohen, E. G., & Catanzarite, L. (1990). Treating status problems in the cooperative classroom. In S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 203-230). New York: Praeger.

Cohen, E. G., Lockheed, M. E., & Lohman, M. R. (1976). The center for interracial cooperation: A field experiment. *Sociology of Education, 49*, 47-58.

Cohen, E. G., & Roper, S. S. (1972). Modification of interracial interaction disability: An application of status characteristic theory. *American Sociological Review, 37*(6), 643-657.

Cohen, E. G., & Sharan, S. (1980). Modifying status relations among Israeli youth. *Journal of Cross-Cultural Psychology, 11*(3), 364-384.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (Rev. ed.).* New York: Academic Press.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist, 45*(12), 1304-1312.

Cohen, P. A., et al. (1982b). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237-248.

Cohen, P. A., Kulik, J. A., & Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.

Coleman, J., Hoffer, T., & Kilgore, S. (1981). *Public and private schools. Final report.* Washington, D.C.: National Center for Education Statistics.

Coleman, J. S. (1961). *The adolescent society*. New York: Free Press.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Congressional Printing Office.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*: Basic Books.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *Public and private high schools: The impact of communities*. New York: Basic Books.

Coley, R. J. (1991). A long track record. *Educational Testing Service Policy News, 4*(1), 1, 7.

Collins, J. (1986). Differential instruction in reading groups. In J. Cook-Gumperz (Ed.), *The social construction of literacy* (pp. 117-137). Cambridge: Cambridge University Press.

Commeyras, M. (1994). Promoting critical thinking through dialogical thinking reading lessons. *Reading Teacher, 46*, 486-494.

Connell, R. W., Ashenden, D. J., Kessler, S., & Dowsett, G. W. (1982). *Making the difference: Schools, families and social division*. Sydney: George Allen & Unwin.

Cook, T. (1974). *Producing equal status interaction between Indian and white boys in British Columbia: An application of expectation training*. Unpublished PhD thesis, Stanford University, CA.

Cook-Gumperz, J. (1986). Introduction: The social construction of literacy. In J. Cook-Gumperz (Ed.), *The social construction of literacy* (pp. 1-15). Cambridge: Cambridge University Press.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

Cooper, H., Valentine, J. C., Nye, B., & Lindsay, J. J. (1999). Relationships between five after-school activities and academic achievement. *Journal of Educational Psychology, 91*, 369-378.

Corno, L. (1989). Self-regulated learning: A volitional analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 111-141). New York: Springer.

Cottle, T. J. (1974). What tracking did to Ollie Taylor. *Social Policy, 5*, 21-24.

Cotton, K. (1996). *Affective and social benefits of small-scale schooling*: ED 401 088.

Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. New York: Cambridge University Press.

Covington, M., & Beery, R. (1976). *Self-worth and school learning*. New York: Holt, Rinehart and Winston.

Covington, M. V. (1987). Achievement motivation, self-attributions and exceptionality. In J. Day & e. a. J. Borkowski (Eds.), *Intelligence and exceptionality: New directions for theory, assessment, and instructional practices* (pp. 173-213). Norwood, NJ: Ablex Publishing.

Crespo, M., & Michelna, J. (1981). Streaming, absenteeism, and dropping out. *Canadian Journal of Education, 6*, 40-55.

Cronbach, L. J., & Snow, R. E. (1969). *Individual differences in learning ability as a function of instructional variables . Final Report*: ED 029 001.

Crosby, M. S., & Owens, E. M. (1991). *An assessment of principal attitudes toward ability grouping in the public schools of South Carolina*: ERIC Document Reproduction Service No. ED364634.

Daiute, C., & Dalton, B. (1993). Collaboration between children learning to write: Can novices be masters? *Cognition and Instruction, 10*, 281-333.

Dale, R. R. (1974). *Mixed or single-sex school?* (Vol. 3). London: Routledge & Kegan Paul.

Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology, 5*(4), 331-343.

Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal, 23*, 357-374.

Dawson, M. M. (1987). Beyond ability grouping: A review of the effectiveness of ability grouping and its alternatives. *School Psychology Review, 16*(3), 25-46.

DeStefano, J., Pepinsky, J., & Sanders, T. (1982). Discourse rules for literacy learning in a first-grade classroom. In L.C. Wilkinson (Ed.), *Communicating in the classroom* . New York: Academic Press.

Dishon, T. J. (1990). The family ecology of boys' peer relations in middle childhood. *Child Development, 61*, 874-892.

Dornbusch, S. M. (1994). *Off the track.* Paper presented at the Biennial Meeting of Sociol. Res. Adolesc., San Diego.

Dreeben, R. (1984). First-grade reading groups: Their formation and change. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 69-84). Orlando, FL: Academic Press.

Dreeben, R., & Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education, 61*, 129-142.

Dreeben, R., & Gamoran, A. (1986). Race, instruction, and learning. *American Sociological Review, 51*, 660-669.

Duffy, G., & Anderson, L. (1981). *Final report: conceptions of reading project.* Unpublished report, Institute for Research on Teaching, Michigan State University, Michigan.

Eccles, J. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75-146). San Francisco: Freeman.

Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 139-186). San Diego: Academic Press.

Eccles, J. S., Wigfield, A., Flanagan, C. A., & Miller, C., et al. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality, 57*(2), 283-310.

Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education, 54*, 151-161.

Eder, D. (1982a). Differences in communicative styles across ability groups. In L. C. Wilkinson (Ed.), *Communicating in the classroom* (pp. 245-264). New York: Academic Press.

Eder, D. (1982b). The impact of management and turn-allocation activities on student performance. *Discourse Processes, 5*, 147-159.

Eder, D. (1983). Family background and ability group assignment. In R. Hauser, D. Mechanic, & A. Haller (Eds.), *Social structure and personality* . New York: Academic Press.

Eder, D. (1995). *School talk: Gender and adolescent culture.* New Brunswick, NJ: Rutgers University Press.

Eder, D., & Felmlee, D. (1984). The development of attention norms in ability groups. In P. Peterson, L. Wilkinson, & N. Hallinan (Eds.), *The social context of instruction* (pp. 189-208). San Diego, CA: Academic Press.

Education Review Office. (1999). *The achievement of boys*. Wellington, NZ: Education Review Office, Ministry of Education.

Eidsomore, R. M. (1964). High school athletes are brighter. *Journal of Health, Physical Education and Recreation, 35*, 53-54.

Eiserman, W. D. (1988). Three types of peer tutoring: Effects on the attitudes of students with learning disabilities and their regular class peers. *Journal of Learning Disabilities, 21*, 249-252.

Elley, W. (1984). To stream or not to stream? *Set, 2*, 1-10.

Elley, W. B., & Irving, J. C. (1985). The Elley-Irving socio-economic index 1981 census revision. *New Zealand Journal of Education Studies, 20*, 115-128.

Emler, N., & Reicher, S. (1995). *Adolescence and delinquency*. Blackwell: Cambridge, MA.

Epple, D., & Romano, R. (1998). Competition between private and public schools vouchers and peer group effects. *American Economic Review, 88*(1), 33-62.

Epstein, J. L. (1983). The influences of friends on achievement and affective outcomes. In J. L. Epstein & N. L. Karweit (Eds.), *Friends in school* (pp. 177-200). New York: Academic Press.

Epstein, J. L., & MacIver, D. J. (1990). *Education in the middle grades: National practices and trends*. Columbus, OH: National Middle School Association.

Erbring, L., & Young, A. A. (1979). Individuals and social structures: Contextual effects as endogenous feedback. *Sociological Methods and Research, 7*, 396-430.

Evans, W. N., Oates, W. E., & Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy, 100*(5), 966-91.

Evans, W. N., & Schwab, R. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics, 110*(4), 941-974.

Evans, W. N., & Schwab, R. M. (1993a). *Finishing high school and stating college: Do Catholic school make a difference?*: University of Maryland, Department of Economics.

Evans, W. N., & Schwab, R. M. (1993b). *Who benefits from private education? Evidence from quartile regressions*: University of Maryland, Department of Economics.

Evertson, C. (1982). Differences in instructional activities in average- and low-achieving junior high English and math classes. *Elementary School Journal, 82*, 329-350.

Evertson, C. M., & Folger, J. K. (1989, March). *Small class, large class: What do teachers do differently?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Felmlee, D., & Eder, D. (1983). Contextual effects in the classroom: The impact of ability groups on student attention. *Sociology of Education, 56*, 77-87.

Felmlee, D., Eder, D., & Tsui, W. (1985). Peer influence on classroom attention. *Social Psychology Quarterly, 48*, 215-226.

Finders, M. (1997). *Just girls: Hidden literacies and life in junior high*. Urbana, IL: National Council of Teachers of English and New York: Teachers College Press.

Finley, M. K. (1984). Teachers and tracking in a comprehensive high school. *Sociology of Education, 57*, 233-243.

Finn, J. D. (1998). *Class size and students at risk: What is known? What is next?* Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, National Institute on the Education of At-Risk Students.

Finn, J. D., & Achilles, C. M. (1990). Answers about questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 557-577.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation & Policy Analysis, 21*(2), 97-109.

Finn, J. D., Folger, J., & Cox, D. (1991). Measuring participation among elementary grade students. *Educational and Psychological Measurement, 51*, 393-402.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal, 95*, 421-434.

Fitz-Gibbon, C. T. (1997). *The value added national project final report*. London: School Curriculum and Assessment Authority.

Floyd, C. (1954). Meeting children's reading needs in the middle grades: A preliminary report. *Elementary School Journal, 55*, 99-103.

Fordham, S., & Ogbu, J. U. (1986). Black students' school success: Coping with the burden of "acting white". *Urban Review, 18*, 176-206.

Fowler, W. J., Jr. (1995). School size and student outcomes. In H. J. Walberg, B. Levin, & J. W. J. Fowler (Eds.), *Organizational Influences on Educational Productivity* (Vol. 5, pp. 3-25). Greenwich, CT: JAI Press.

Freedman, S. W. (Ed.). (1985). *The acquisition of written language: Response and revision*.

Frentz, C., Cresham, F. M., & Elliots, S. N. (1991). Popular, controversial, neglected, and rejected adolescents: Contrasts of social competence and achievement differences. *Journal of School Psychology, 29*, 109-120.

Friedkin, N., & Necochea, J. (1988). School system size and performance: A contingency perspective. *Education Evaluation and Policy Analysis, 10*(3), 237-249.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997a). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*, 174-206.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Philips, N. B., Karns, K., & Dutke, S. (1997b). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *Elementary School Journal, 97*(3), 223-249.

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal, 99*(3), 201-.

Fuchs, L. S., Fuchs, D. H., Hamlet, C. L., & Karns, K. (1998). High-achieving students interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35*(2), 227-287.

Fulwood, S. (1991, January 1,). Attitudes on minorities in conflict. *Los Angeles Times,*, pp. 1, sec. A.

Fung, I. Y. Y. (1999). *L1-assisted reciprocal teaching for ESL students.* Unpublished Masters Thesis, University of Auckland, Auckland, NZ.

Gaffney, M., & Smith, A. (1999). *Evaluation of the TIE Scheme: Summary Report.* Dunedin, NZ: University of Otago.

Gambrell, L. (1984). How much time do children spend reading during teacher-directed reading instruction? In J. Niles & L. Harris (Eds.), *Changing perspectives on research in reading/language processing and instruction (Thirty-Third Yearbook of the National Reading Conference)* . Rochester, NY: National Reading Conference.

Gambrell, L. B., Wilson, R. M., & Gantt, W. N. (1981). Classroom observations of task-attending behaviors of good and poor readers. *Journal of Educational Research, 74*, 400-404.

Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education, 59*(4), 185-98.

Gamoran, A. (1987a). The stratification of high school learning opportunities. *Sociology of Education, 60*, 135-55.

Gamoran, A. (1987b). Organization, instruction, and the effects of ability grouping: Comment on Slavin's "best-evidence synthesis". *Review of Educational Research, 57*(3), 341-345.

Gamoran, A. (1989). Measuring curriculum differentiation. *American Journal of Education, 97*, 129-143.

Gamoran, A. (1992). Social factors in education. In M. C. Alkin (Ed.), *Encyclopaedia of Educational Research* (pp. 1222-1229). New York: Macmillan.

Gamoran, A. (1993). Alternative uses of ability grouping in secondary schools: Can we bring high-quality instruction to low-ability classes? *American Journal of Education, 102*(1), 1-22.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research, 57*, 415-435.

Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94*(5), 1146-83.

Gamoran, A., Nystrand, M., Berends, M., & LePore, P. C. (1995). An organizational analysis of the effects of ability grouping. *American Educational Research Journal, 32*, 687-715.

Garcia, E. (1991). *Education of linguistically and culturally diverse students: Effective instructional practices. Report 1*: ERIC Document ED 338 099.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than p values: Estimation rather than hypothesis testing. *Statistics in Medicine, 292*, 746-750.

Garet, M. S., & Delany, B. (1988). Students, courses and stratification. *Sociology of Education, 61*, 61-77.

Geen, R. G. (1991). Social motivation. In M. R. Rosenzweig, L. W. Porter, & et al. (Eds.), *Annual review of psychology* (Vol. 42, pp. 377-399). Palo Alto, CA: Annual Reviews, Inc.

George, P. S., & Shewey, K. (1994). *New evidence for the middle school*. Columbus, OH: National Middle School Association. ED 396 839.

Gerber, S. B. (1996). Extracurricular activities and academic achievement. *Journal of Research and Development in Education, 30*, 42-50.

Gibson, A., & Asthana, S. (1998). School performance, school effectiveness and the 1997 White Paper. *Oxford Review of Education, 24*(2), 195-210.

Gillibrand, E., Robinson, P., Brawn, R., & Osborn, A. (1999). Girls' participation in physics in single sex classes in mixed schools in relation to confidence and achievement. *International Journal of Education, 21*(4), 349-362.

Gilroy, A., & Moore, D. W. (1988). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities with ten primary school girls. *Educational Psychology: An International Journal of Experimental Educational Psychology, 8*, 41-49.

Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly, 12*(2), 134-149.

Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School size: Research and policy*. Beverly Hills, CA: Sage.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverley Hills, CA: Sage Publications.

Glass, G. V., & Smith, M. J. (1978). *Meta-analysis of research on the relationship of class-size and achievement*: Laboratory of Educational Research, University of Colorado.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation & Policy Analysis, 1*(1), 2-16.

Glass, S. R. (1997). Markets and myths: Autonomy in public and private schools. *Education Policy Analysis Archives, 5*(1), (http://olam.ed.asu.edu/epaa).

Goldhaber, D. (1996). Public and private high schools: Is school choice an answer to the productivity problem? *Economics of Education Review, 15*(2), 93-109.

Goldring, E. B. (1990). Assessing the status of information on classroom organizational frameworks for gifted students. *Journal of Educational Research, 83*(6), 313-326.

Good, T. I. (1981). Teacher expectations and student perception: A decade of research. *Educational Leadership, 38*, 415-423.

Good, T. L., & Beckerman, T. M. (1978). Time on task: A naturalistic study in sixth grade classrooms. *Elementary School Journal, 78*, 192-201.

Goodlad, J. I. (1984). *A place called school*. New York: McGraw-Hill.

Goodlad, J. I., & Anderson, R. H. (1987). *The nongraded elementary school*. New York: Teachers College Press (Original work published 1963).

Grabe, M. (1976). Big school, small school: Impact of the high school environment. *Contemporary Educational Psychology, 1*, 20-15.

Grabe, M. (1981). School size and the importance of school activities. *Adolescence, 61*, 21-31.

Grant, K., & Rothenberg, J. (1981, April). *Charting educational futures: Interaction patterns in first and second grade reading groups.* Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Grant, L. (1986). *Classroom peer relationships of minority and nonminority students.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Gray, J., Jesson, D., & Sime, N. (1990). Estimating differences in the examination performances of secondary schools in six LEAs: A multi-level approach to school effectiveness. *Oxford Review of Education, 16*, 137-156.

Guay, F., Boivin, M., & Hodges, V. E. (1999). Predicting change in academic achievement: A model of peer experiences and self-system process. *Journal of Educational Psychology, 91*, 105-115.

Gump, P. V., & Friesen, W. V. (1964). Satisfactions derived from nonclass settings. In R. G. Barker & P. V. Gump (Eds.), *Big school, small school: High school size and student behavior* (pp. 75-93). Stanford, CA: Stanford University Press.

Guntermann, E., & Tovar, M. (1987). Collaborative problem-solving with Logo: Effects of group size and group composition. *Journal of Educational Computing Research, 3*, 313-334.

Gutierrez, R. O., & Slavin, R. E. (1992, June). *Achievement effects of the nongraded elementary school: A retrospective review*: ED 346 996.

Haller, E. J., & Davis, S. A. (1980). Does socioeconomic status bias the assignment of elementary school students to reading groups? *American Educational Research Journal, 17*(4), 409-418.

Hallinan, M. (1984). Summary and implications. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 229-240). Orlando, FL: Academic Press.

Hallinan, M. T., & Williams, R. A. (1989). Interracial friendship choices in secondary schools. *American Sociological Review, 54*, 67-78.

Hallinger, P., & Murphy, J. F. (1986). The social context of effective schools. *American Journal of Education, 94*, 328-355.

Hanks, M. P., & Eckland, B. K. (1976). Athletics and social participation in the educational attainment process. *Sociology of Education, 49*, 271-294.

Hanushek, E. (1986a). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(1141-1177).

Hanushek, E. A. (1986b). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-77.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation & Policy Analysis, 19*(2), 141-164.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation & Policy Analysis, 21*(2), 143-163.

Hanushek, E. A., Rivkin, S. G., & Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics & Statistics, 78*(4), 611-627.

Hargreaves, D. H. (1967). *Social relations in a secondary school*. London: Tinling.

Hargreaves, L., Galton, M., & Pell, A. (1998). The effects of changes in class size on teacher-pupil interaction. *International Journal of Educational Research, 29*.

Harker, R., & Nash, R. (1996). Academic outcomes and school effectiveness: Type A and type B effects. *New Zealand Journal of Educational Studies, 31*, 143-170.

Harrah, D. D. (1956). A study of the effectiveness of five kinds of grouping in the classroom. *Dissertation Abstracts International, 17*, 715 (University Microfilms No. 56-1114).

Harris, J. R. (1995). Where is the child's environment? A group socialization theory of development. *Psychological Review, 102*(3), 458-89.

Hart, S. (1982). Analyzing the social organization for reading in one elementary school. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* . New York: Holt, Rinehart & Winston.

Harter, S. (1990). Self and identity development. In S. S. Feldman & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 352-387). Cambridge, MA: Harvard University Press.

Hartup, W. W. (1998). The company they keep: Friendships and their developmental significance. In E. E. Hertzig & e. a. E. A. Farber (Eds.), *Annual progress in child psychiatry and child development: 1997* (pp. 63-77). Bristol, PA: Brunner/Mazel.

Hartup, W. W. (1999). Constraints on peer socialization: Let me count the ways. *Merrill-Palmer Quarterly-Journal of Developmental Psychology, 45*(1), 172-183.

Haskins, R., Walden, T., & Ramey, C. T. (1983). Teacher and student behavior in high- and low-ability groups. *Journal of Educational Psychology, 75*, 865-876.

Hattie, J. (1990). Performance indicators in education. *Australian Journal of Education, 34*(3), 249-276.

Hattie, J. (1992). Measuring the Effects of Schooling. *Australian Journal of Education, 36*(1), 5-13.

Heap, J. L. (1986). *Collaborative practices during computer writing in a first grade classroom.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Heath, A. F., & Blakey, L. S. (1992). Differences between comprehensive schools: Some preliminary findings. In D. Reynolds & P. Cuttance (Eds.), *School Effectiveness: Research policy and practice* . London: Cassell.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Stock, W. (1983). The effects of class size: An examination of rival hypotheses. *American Educational Research Journal, 20,* 63-65.

Hertz-Lazarowitz, R., Kirkus, V. B., & Miller, N. (1995). Implications of current research on cooperative interaction for classroom application. In R. Hertz-Lazarowitz & e. a. N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning.* (Vol. viii, pp. 294). New York, NY: Cambridge University Press.

Heyns, B. (1986). Educational effects: Issues in conceptualization and measurement. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 305-340). New York: Greenwood.

Hicks, L., & Townsend, M. A. R. (1993). *Social satisfaction and the perceived value of academic tasks.* Paper presented at the annual meeting of the American Psychological Association, Toronto.

Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. *Reading Research Quarterly, 18,* 231-255.

Ho, E., & Willms, J. D. (1996). Effects of parental involvement on eighth-grade achievement. *Sociology of Education, 69,* 126-141.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14*(3), 205-27.

Hoffman, J. V., O'Neal, S. F., Kastler, L. A., Clements, R. O., Segel, K. W., & Nash, M. F. (1984). Guided oral reading and miscue focused verbal feedback in second-grade classrooms. *Reading Research Quarterly, 19367-384.*

Hogue, A., & Steinberg, L. (1995). Homophily of internalized distress in adolescent peer groups. *Developmental Psychology, 31,* 897-906.

Holland, A., & Andre, T. (1987). Participation in extracurricular activities in secondary school: What is known, what needs to be known? *Review of Educational Research, 57*(4), 437-466.

Howley, C. (1995). The matthew principle: A West Virginia replication? *Education Policy Analysis Archieves, 3*(18), (http://epaa.asu.educ/epaa/v3n18.html).

Huguet, P., & Monteil, J. (1995). The influence of social comparison with less fortunate others on task performance: The role of gender motivations or appropriate norms. *Sex Roles, 33,* 753-765.

Hunter, D. (1978). *Student on-task behavior during reading group meeting.* Unpublished dissertation, University of Missouri, Columbia, MO.

Hymel, S., Comfort, C., Schonert-Reichl, K., & McDougall, P. (1996). Academic failure and school dropout: The influence of peers. In J. Juvonen & K. R. Wentzel (Eds.), *Social motivation: Understanding children's school adjustment* (pp. 313-345). New York: Cambridge University Press.

Ide, J. K., Parkerson, J., Haertel, G. D., & Walberg, H. J. (1981). Peer group influence on educational outcomes: A quantitative synthesis. *Journal of Educational Psychology, 73,* 472-484.

Imai, M., Anderson, R. C., Wilkinson, I. A. G., & Yi, H. (1992). Properties of attention during reading lessons. *Journal of Educational Psychology, 84,* 160-173.

Ingels, S. J., et al. (1988). *Symposium on the National Education Longitudinal Study of 1988 (NELS:88) and the NELS:88 Field Test (New Orleans, Louisiana, April 5-9, 1988)*: ED 297 006.

Jacques, N., Wilton, k. M., & Townsend, M. A. R. (1998). Cooperative learning and social acceptance of children with mild intellectual disability. *Journal of Intellectual Disability Research, 42*, 29-36.

Jaeger, R. M., & Hattie, J. A. (1996). Artifact and artiface in education policy analysis: It's not all in the data. *School Administrator, 53*, 24-25, 28-29.

Jason, L. A., Erone, L., & Soucy, G. (1979). Teaching peer tutoring behaviors in first- and third-grade classrooms. *Psychology in the Schools, 16*, 261-269.

Jencks, C. (1985, April). How much do high school students learn? *Sociology of Education, 58*, 128-135.

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In J. Laurence, E. Lynn, & M. G. H. McGeary (Eds.), *Inner-city poverty in the United States* . Washington, D.C.: National Academy Press.

Johnson, D., Johnson, R., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. *Review of Educational Research, 53*, 5-54.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, Min: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (1995). *Social interdependence: Cooperative learning in education*. San Francisco: Jossey-Bass.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89*(1), 47-62.

Johnston, J. H., & Markle, G. C. (1983). What research says to the practitioner. *Middle School Journal (National Middle School Association), 14*(4), 53-57.

Joiner, R., Messer, D., Light, P., & Littleton, K. (1998). The effects of gender, expectations of success and social comparison on children's performance on a computer-based task. *Educational Psychology, 18*, 319-326.

Jones, D. M. (1948). An experiment in adaptation to individual differences. *Journal of Educational Pyschology, 39*(5), 257-272.

Jones, J. C., Shallcrass, J., & Dennis, C. C. (1972). Co-education and adolescent values. *New Zealand Journal of Educational Studies, 30*(2), 334-341.

Juel, C. (1990). Effects of reading group assignment on reading development in first and second grade. *Journal of Reading Behavior, 22*, 233-254.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31*(3), 268-289.

Kaiden, E. (1994). Repeat attendance at a college tutoring center: The students' perspectives. *Research and Teaching in Developmental Education, 11*, 49-62.

Karabenick, S. A., & Sharma, R. (1994a). Perceived teacher support of student questioning in the college classroom: Its relation to student characteristics and role in the classroom questioning process. *Journal of Educational Psychology, 86*(1), 90-103.

Karabenick, S. A., & Sharma, R. (1994b). Seeking academic assistance as a strategic learning resource. In P. R. Pintrich & D. R. Brown, et al. (Eds.), *Student motivation, cognition, and learning: Essays in honor of Wilbert J. McKeachie* (pp. 189-211). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kealey, M. (1984). Meats and veges: An ethnographic study of two grammar school classes. *Set, 2*, 10.

Kelly, M., Moore, D. W., & Tuck, B. F. (1994). Reciprocal Teaching in a Regular Primary School Classroom. *Journal of Educational Research, 88*(1), 53-61.

Kenway, J., & Willis, S. (1986). Feminist single-sex strategies: Some theoretical flaws and practical fallacies. *Discourse, 7*(1), 1-30.

Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review, 51*, 842-858.

Kerslake, J., Lange, D., & Bennie, N. (1998, May). A summary report of the 1996 truancy action survey. In S. Kennedy (Ed.), *The Research Bulletin, No.8* (pp. 53-59). Wellington, NZ: Ministry of Education.

Kerslake, J., Murrow, K. E., & Lange, D. (1998). Exempt from school, but learning at home: An exploratory study of home schooling in Aotearoa New Zealand. In M. o. Education (Ed.), *The Research Bulletin, No. 8*. Wellington, NZ: Ministry of Education.

Kindermann, T. A. (1993). Natural peer groups as contexts for individual development: The case of children's motivation in school. *Developmental Psychology, 29*(6), 970-977.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*, 664-687.

King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist, 27*, 111-126.

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90*(1), 1314-152.

Kiser-Kling, K. (1995). *Life in a small teacher-pupil ration class.* Unpublished Ed. D dissertation, University of North Carolina, Greensboro.

Knight, G. P., & Bohlmeyer, E. M. (1990). Cooperative learning and achievement: Methods for assessing causal mechanisms. In e. a. S. Sharan (Ed.), *Cooperative learning: Theory and research* (pp. 1-22). New York,NY: Praeger Publishers.

Knight, S., & Stallings, J. (1995). The implementation of the Accelerated School model in an urban elementary school. In R. Allington & s. Walmsley (Eds.), *No quick fix: Rethinking literacy programs in American elementary schools* (pp. 236-252). New York: Teachers College Press.

Knowles, J. G. (1998). Home education: Personal histories. In M. L. Fuller & G. Olsen (Eds.), *Home-school relations: Working successfully with parents and families* (pp. 302-331). Toronto: Allyn and Bacon.

Krappman, L. (1985). The structure of peer relationships and possible effects on school achievement. In R. A. Hinde, A. N. Perret-Clermont, & J. Stevenson-Hinde (Eds.), *Social relationships and cognitive development* (pp. 149-166). Oxford, England: Clarendon Press.

Krueger, A. B. (1997, May). *Experimental estimates of education production functions* (Working paper #6051). Cambridge, MA: National Bureau of Economics Research.

Kruse, A.-M. (1987a). Piger for sig, sammen. *Kvinden og Samfundet*, 5.

Kruse, A.-M. (1987b). *Sagde du konssegregering -- med vilje? Paedagogik med rode stromper*. Kobenhavn: Danmarks Laererhojskole.

Kruse, A.-M. (1989a). Hvorfor pigeklasser? In Adda Hilden & A.-M. Kruse (Eds.), *Pigernes skole* (pp. 249-263). Skive: Klim.

Kruse, A.-M. (1989b). Pigeskile eller faellesskile? Debatten omkring arhundredeskiffet. In Adda Hilden & A.-M. Kruse (Eds.), *Pigernes skole* (pp. 43-77). Skive: Klim.

Kruse, A.-M. (1990). Konsadskilt undervisning som konsbevidst paedagogik. In Helle Jacobsen & L. Hojgaard (Eds.), *Skolen er kon* (pp. 36-81). Viborg: Ligestillingsridet.

Kruse, A.-M. (1991). Metoder til fremme af ligevaerd. In Gudrun Pedersen & K. Reisby (Eds.), *Ligevard -- Mangfoldighed* (pp. 66-80). Viborg: Danmarks Laererhojskole.

Kruse, A.-M. (1992). We have learnt not just to sit back, twiddle our thumbs and let them take over. Single-sex settings and the development of a pedagogy for girls and a pedagogy for boys in Danish schools. *Gender and Eduacation, 4*(12), 81-103.

Kruse, A.-M. (1994a). Er drenge er bare drenge? *Tidsskrift for borne & ungdomskultur, 34*, 65-79.

Kruse, A.-M. (1994b). Hvordan er det med der forskelle pa piger og drenge? Interview med Harriet Bjerrum Nielsen. *Tidsskrift for borne & ungdomskultur, 34*, 51-65.

Kruse, A.-M. (1994c). *Hvorfor konsopdelt paedagogik? Om polariserende paedogogik som metode til at modvirke ulighed mellem konnene belyst via Projekt Pigeklasse-Drengeklasse*. Aarhus: Denmark.

Kruse, A.-M. (1994d). Kommunikation og dominans i klasserrummet. *CekvinaNyt*, 2.

Kruse, A.-M. (1995, January 10-12). *Single-sex settings: Pedagogies for girls and boys in Danish schools.* Paper presented at the UNESCO Colloquium, Institute of Education, London.

Kruse, A.-M. (1996). Approaches to teaching girls and boys. Current debates, practices, and perspectives in Denmark. *Women's Studies International Forum, 19*(4), 429-445.

Kulik, C.-L. C. (1985). *Effects of inter-class ability grouping on achievement and self esteem.* Paper presented at the Annual Meeting of the American Psychological Association, Los Angeles.

Kulik, C.-L. C., & Kulik, J. A. (1982a). Effects of ability grouping on secondary school students: A meta-analysis. *American Educational Research Journal, 19*(3), 415-428.

Kulik, C.-L. C., & Kulik, J. A. (1982b). Research synthesis on ability grouping. *Educational Leadership, 39*, 619-621.

Kulik, C.-L. C., & Kulik, J. A. (1984). *Effects of ability grouping on elementary school pupils: A meta-analysis.* Paper presented at the Annual Meeting of the American Psychological Association, Toronto, Ontario, Canada.

Kulik, J. A. (1993). *An analysis of the research on ability grouping*: ED 367 095.

Kulik, J. A., & Kulik, C.-L. C. (1985). *Effects of inter-class ability grouping on achievement and self-esteem.* Paper presented at the Annual Convention of the American Psychological Association (93rd), Los Angeles, CA.

Kulik, J. A., & Kulik, C.-L. C. (1987). Effects of ability grouping on student achievement. *Equity and Excellence, 23*, 22-30.

Kulik, J. A., & Kulik, C.-L. C. (1990). Ability grouping and gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 178-196). Boston: Allyn & Bacon.

Kulik, J. A., & Kulik, C.-L. C. (1991). *Research on ability grouping: Historical and contemporary perspectives.* Storrs: University of Connecticut, National Research Center on the Gifted and Talented. (ERIC Document Reproduction Service No. ED 350 777).

Kulik, J. A., & Kulik, C.-L. C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36*(2), 73-77.

Kutnick, P. (1994). Does preschool curriculum make a difference in primary school performance: Insights into the variety of preschool activities and their effects on school achievement and behavior in the Caribbean Island of Trinadad. *Early Childhood Development & Care, 103*, 27-42.

Ladd, G. W. (1990). Having friends, Keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61*(4), 1081-1100.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development, 67*(3), 1103-1118.

Lamborn, S. D., Brown, B. B., Mounts, N. S., & Steinberg, L. (1992). Putting school in perspective: The influence of family, peers, extracurricular participation, and part-time work on academic engagement. In F. M. Newman (Ed.), *Student engagement and achievement in American secondary schools* (pp. 153-181). New York: Teachers College Press.

Lareau, A. (1987). Social class differences in family school relationships: The importance of cultural capital. *Sociology of Education, 60*, 73-85.

Lareau, A. (1989). *Home advantage. Philadelphia.* Philadelphia: Falmer Press.

Larnen, S. C., & Ehly, S. (1976). Peer tutoring: An aid to individual instructions. *The Clearing House, 49*, 273-277.

Latane, B., Willianms, K. D., & Harkins, S. G. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822-832.

Lauder, H., & Hughes, D. (1990). Social inequalities and differences in school outcomes. *New Zealand Journal of Educational Studies, 25*, 37-60.

Lauder, H., Hughes, D., Waslander, S., Thrupp, M., McGlinn, J., Newton, S., & Dupuis, A. (1994). *The creation of market competition for education in New Zealand*. Wellington, NZ: Ministry of Education.

Lauder, H., Hughes, D., Watson, D., Waslander, S., Thrupp, M., Strathdee, R., Simiyu, I., Dupuis, A., McGlinn, J., & Hamlin, J. (1999). *Trading in futures: Why markets in education don't work*. Buckingham; Philadelphia: Open University Press.

Le Fevre, D. (1996). *Tape-assisted reciprocal teaching for readers with poor decoding skills.* Unpublished Masters Thesis, The University of Auckland, Auckland.

Lee, V. E., & Bryk, A. S. (1986). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 78*, 381-395.

Lee, V. E., & Bryk, A. S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education, 61*, 78-94.

Lee, V. E., & Bryk, A. S. (1989a). A multilevel model of the social distribution of high school achievement. *Sociology of Education, 62*, 172-192.

Lee, V. E., & Bryk, A. S. (1996). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology, 78*, 381-395.

Lee, V. E., Bryk, A. S., & Smith, J. B. (1993). The effects of high school organization on teachers and students. In L. Darling-Hammond (Ed.), *Review of Research in Education* (Vol. 19, pp. 171-268). Washington, DC: American Educational Research Association.

Lee, V. E., & Croninger, R. G. (1994). The relative importance of home and school in the development of literacy skills for middle-grade students. *American Journal of Education, 102*, 286-329.

Lee, V. E., & Smith, J. B. (1993). Effects of school restructuring on the achievement and engagement of middle-grade students. *Sociology of Education, 60*(3), 164-187.

Lee, V. E., & Smith, J. B. (1995). Effects of high school restructuring and size on gains in the achievement and engagement for early secondary school students. *Sociology of Education, 68*(4), 241-270.

Lee, V. E., & Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis, 19*(3), 205-227.

LePore, P. C., & Warren, J. R. (1997). A comparison of single-sex and coeducational Catholic secondary schooling: Evidence from the national educational longitudinal study of 1988. *American Educational Research Journal, 34*(3), 485-511.

Levin, H. M. (1996). Accelerated schools after eight years. In L. Schauble & R. Glaser (Eds.), *Innovations in learning new environments for education* (pp. 329-352). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Levin, H. M. (1998). Educational vouchers: Effectiveness, choice, and costs. *Journal of Policy Analysis and Management, 17*(3), 373-392.

Levin, H. M., Glass, G. V., & Meister, G. R. (1984). *Cost-effectiveness of four educational interventions*. California: Institute for Research on Educational Finance and Governance, Stanford University. (Eric Reproduction Service No. ED 246 533).

Levine, J. M., & Moreland, R. L. (1990). Progess in small group research. *Annual Review of Psychology, 41*, 585-634.

Levine, J. M., Resnick, L. B., & Higgins, E. T. (1993). Social foundations of cognition. *Annual Review of Psychology, 44*, 585-612.

Light, P., Littleton, K., Messer, D., & Joiner, R. (1994). Social and communicative processes in computer-based problem solving. *European Journal of Psychology of Education, 11*(1), 93-109.

Lindle, J. C. (1994). *Review of the literature on tracking and ability grouping. Second draft*: ED 384 643.

Link, C. R., & Mulligan, J. G. (1991). Classmates' effects on black student achievement in public school classrooms. *Economics of Education Review, 10*(4), 297-310.

Lockheed, M. E., & Hall, K. (1976). Conceptualizing sex as status characteristics: Applications to leadership training strategies. *Journal of Social Issues, 32*, 111-124.

Lockheed, M. E., Harris, A. M., & Nemceff, W. P. (1983). Sex and social influence: Does sex function as a status characteristic in mixed-sex groups of children? *Journal of Educational Psychology, 75*, 877-888.

Lockwood, A. T. (1990). Mixed reviews: Two interviews. *National Center on Effective Secondary School Newsletter, 5*(1), 9-11.

Lockwood, A. T. (1996). *Tracking: Conflicts and resolutions. Controversial issues in education*. Thousand Oaks, CA: Corwin Press.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*, 423-458.

Lounsbury, J. H., & Clark, D. C. (1990). *Inside grade eight: From apathy to excitement*. Reston, VA: National Association of Secondary School Principals. ED 327 318.

Loveless, T. (1993). *Organizational coupling and the implication of tracking reform* (Technical Report No.143). Chicago, IL: Midwest Administration Center, the University of Chicago.

Loveless, T. (1998). *The tracking and ability grouping debate. Volume 2, Number 8*. Washington, DC: Thomas B. Fordham Foundation, (http://www.edexcellence.net).

Loveless, T. (1999a). *The tracking and ability grouping*. Washington D.C.: Thomas B. Fordham Foundation, Washington, D.C. (http://www.edexcellence.net) ERIC Document Reproduction Service No. ED 422 454.

Loveless, T. (1999b). *The tracking war: State reform meets school policy*. Washington, D.C.: Bookings Institution Press.

Mahard, R. E., & Crain, R. L. (1983). Research on minority achievement in desegregated schools. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* . Philadelphia: Temple University Press.

Mahoney, J. L., & Cairns, R. B. (1997). Do extracurricular activities protect against early school dropout? *Developmental Psychology-, 33*(2), 241-253.

Malloy, L., & Gilman, D. (1989). The cumulative effects on basic skills achievement of Indiana's PEIME TIME. A state sponsored program of reduced class size. *Contemporary Education, 60*, 169-172.

Marsh, H. W. (1984a). Self-concept, social comparison and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal, 21*, 799-806.

Marsh, H. W. (1984b). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education, 28*, 165-181.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*, 280-295.

Marsh, H. W. (1989). Effects of attending single-sex and coeducational high schools on achievement, attitudes, behaviors, and sex differences. *Journal of Educational Psychology, 81*, 70-85.

Marsh, H. W. (1992). Extracurricular activities: Beneficial extension of traditional curriculum or subversion of academic goals? *Journal of Educational Psychology, 84*, 553-562.

Marsh, H. W., & Parker, J. W. (1984). Determinants of self-concept: Is it better to be a relatively large fish in s small pond even if you don't learn to swim as well. *Journal of Personality and Social Psychology, 47*, 213-231.

Marsh, H. W., & Rowe, K., J. (1996). The effects of single-sex and mixed-sex mathematics classes within a coeducational school: A reanalysis and comment. *Australian Journal of Education, 40*(2), 147-162.

Martin, J., & Evertson, C. M. (1980). *Interactions with reading groups of differing ability levels.* Austin: University of Texas, Research & Development Center for Teacher Education (ERIC Doc. Rep. Service No. ED 203 303).

Mason, D. A., & Burns, R. B. (1995). Teachers' views of combination classes. *Journal of Educational Research, 89*(1), 36-45.

Mason, D. A., & Burns, R. B. (1996). "Simply no worse and simply no better" may simply be wrong: A critique of Veeman's conclusion about multigrade classes. *Review of Educational Research, 66*(3), 307-322.

Mason, D. A., & Burns, R. B. (1997a). Reassessing the effects of combination classes. *Educational Research and Evaluation, 3*(1), 1-53.

Mason, D. A., & Burns, R. B. (1997b). Toward a theory of combination classes. *Educational Research & Evaluation (An International Journal on Theory & Practice), 3*(4), 281-304.

Mason, D. A., & Burns, R. B. (1998). *Assignment of teachers to combination classes*: Unpublished paper.

Mason, D. A., & Doepner, R. W. (1998). Principals' views of combination classes. *Journal of Educational Research, 91*(3), 160-172.

Mason, D. A., & Good, T. L. (1996). Mathematics instruction in combination and single-grade classes: An exploratory investigation. *Teachers College Record, 98*(2), 236-265.

Mason, D. A., Schroeter, D. D., Combs, R. K., & Washington, K. (1992). Assigning average-achieving eighth graders to advanced mathematics classes in urban junior high. *Elementary School Journal, 92*(5), 587-599.

Mason, L. (1998). Sharing cognition to construct scientific knowledge in school context: The role of oral and written discourse. *Instructional Science, 26*, 359-389.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33*(1), 62-94.

Maughan, B., & Rutter, M. (1987). Pupils progress in selective and non-selective schools. *School Organisation, 7*, 49-68.

Mayberry, M., Knowles, J. G., Ray, B., & Marlow, S. (1995). *Home schooling: Parents as educators*. Thousand Oaks: Corwin Press.

McAuliffe, T., Dembo, J., & Myron, H. (1994). Status rules of behaviour in scenarios of peer learning. *Journal of Educational Psychology, 86*(2), 163-172.

McCarthey, S. J., & McMahon, S. (1992). From convention to invention: Three approaches to peer interactions during writing. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups* . Cambridge: Cambridge University Press.

McCarthy, J., & Still, S. (1993). Hollibrook accelerated elementary school. In J. Murphy & P. Hallinger (Eds.), *Restructuring schools: Learning from ongoing efforts* (pp. 63-83). Monerey Park, CA: Corwin Press.

McCaslin, M., & Good, T. L. (1996). The informal curriculum. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 622-670). New York: Macmillan.

McDermott, R. P. (1976). *Kids make sense: An ethnographic account of the international management of success and failure in one first-grade classroom.* Unpublished doctoral dissertation, Stanford University.

McFadden, A. C., Marsh, G. E., Price, B. J., & Hwang, Y. (1992). A study of race and gender bias in the punishment of school children. *Education and Treatment of Children, 15*, 140-146.

McGiverin, J., Gilman, D., & Tillitski, C. (1989). A meta-analysis of the relation between class size and achievement. *Elementary School Journal, 90*(1), 47-56.

McNaughton, S. (1995). *Patterns of emergent literacy: Processes of development and transition*. Melbourne: Oxford University Press.

McNeal, R. B. J. (1999). Participation in high school extracurricular activities: Investigating school effects. *Social Science Quarterly, 80*(2), 291-309.

McPherson, A., & Willms, J. D. (1987). Equalisation and improvement: Some effects of comprehensive reorganisation in Scotland. *Sociology, 21*, 509-539.

Meece, J. L., Wiggield, A., & Eccles, J. S. (1990). Predictors of math anxiety and its consequences for young adolescents' course enrollment intentions and performances in mathematics. *Journal of Educational Psychology, 82*, 60-70.

Mehan, H. (1972). "What time is it, Denise?": Asking known information questions classroom discourse. *Theory into Practice, 18*(4), 285-294.

Meighan, R. (1995). Home-based education effectiveness research and some of its implications. *Educational Review, 47*(3), 275-287.

Metz, M. (1978). *Classrooms and corridors: The crisis of authority in desegregated secondary schools*. Berkely: University of California Press.

Metz, M. H. (1990). How social class differences shape teachers work. In M. W. McLaughlin, J. E. Talbert, & N. Bascia (Eds.), *The contexts of teaching in secondary schools* . New York: Teachers College Press.

Milligan, S., & Thomson, K. (1992). *Listening to girls*. Melbourne: Curriculum Corporation.

Mills, R. (1998). *Grouping Students for Instruction in Middle Schools*. Champaign, IL: Eric Clearinghouse (http://www.ed.gov/database/Eric_Digests/ed419631.html).

Ministry of Education. (1991). *Report of the economic and educational viability of small school review*. Wellington, NZ: Ministry of Education.

Ministry of Education. (1996). *New Zealand schools/Nga Kura o Aotearoa statistical annex 1995*. Wellington, NZ: Ministry of Education.

Ministry of Education. (1999a). *Education statistics of New Zealand for 1998*. Wellington, NZ: The Data Management and Analysis Division, Ministry of Education.

Ministry of Education. (1999b). *Small primary schools*. Wellington, NZ: Ministry of Education, Education Review Office.

Ministry of Education. (1999c). *New Zealand schools, Nga Kua o Aotearoa 98: A report on the compulsory schools sector in New Zealand*. Wellington, NZ: Ministry of Education.

Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation & Policy Analysis, 21*(2), 165-177.

Moreland, R. L., & Levine, J. M. (1992). The composition of small groups. *Advances in Group Processes, 9*, 237-80.

Mortimore, P. (1995). The balancing act. *Education Guardian*.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters*. Berkeley, Calif.: University of California Press.

Mosteller, F., Light, R. J., & Sachs, J. (1996). Sustained inquiry in education: Lessons from skill grouping and class size. *Harvard Educational Review, 66*(4), 797-842.

Mpofu, E. (1997). Children's social acceptance and academic achievement in Zimbabwean multicultural school settings. *Journal of Genetic Psychology, 158*(1), 5-24.

Mueller, D. J.. Chase, C. I., & Walden, J. D. (1988). Effects of reduced class size in primary classes. *Educational Leadership, 45*(7), 48-50.

Nash, R., & Harker, R. (1998). *Making progress: adding value in secondary education*. Palmerston North, NZ: Massey University Printery.

Nash, R., & Harker, R. K. (1997). *Progress at school: Final report to the Ministry of Education.* Wellington, NZ: Ministry of Education.

National Center for Educational Statistics. (1985). *High school and beyond: An analysis of course-taking patterns in secondary schools as related to student characteristics.* Washington, DC: U.S. Government Printing Office.

Newcomb, A. F., & Bagwell, C. L. (1995). Children's friendship relations: A meta-analytic review. *Psychological Bulletin, 117*, 306-347.

Newmann, F., & Thompson, J. A. (1987). *Effects of cooperative learning on achievement in secondary schools: A summary of research.* Madison, WI: University of Wisconsin-Madison, National Center on Effective Secondary Schools.

Nguyen-Jahiel, K., Anderson, R. C., Waggoner, M., & Rowell, B. (1998). From recitation to collaborative reasoning: A chronicle of one teacher and her students. In R. Horowitz (Ed.), *The evolution of talk about text* . Newark,DE: International Reading Association.

Noland, T. K., & Taylor, B. L. (1986). *The effects of ability grouping: A meta-analysis of research findings.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Nuthall, G. A., & Alton-Lee, A. G. (1990). Research on teaching and learning: Thirty years of change. *The Elementary School Journal, 90*(5), 547-570.

Nuthall, G. A., & Alton-Lee, A. G. (1992). Understanding how students learn in classroom. In M. Pressley, K. R. Harris, & J. T. Guthrie (Eds.), *Promoting academic competence and literacy in school* . San Diego, California: Academic Press.

Nuthall, G. A., & Alton-Lee, A. G. (1993). Predicting learning from student experience of teaching: A theory of student knowledge construction in classroom. *American Educational Research Journal, 30*(4), 799-840.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation & Policy Analysis, 21*(2), 127-142.

Nystrand, M., & Gamoran, A. (1988). *A study of instruction as discourse.* Madison, WI: Wisconsin Center for Education Research.

Oakes, J. (1981). *A question of access: Tracking and curriculum differentiation in a national sample of English and mathematics classes. A study of schooling in the United States*: Technical Report Series, No. 24 ED 214 892.

Oakes, J. (1982). Classroom social relationships: Exploring the Bowles and Gintis hypothesis. *Sociology of Education, 55*, 197-212.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven, CT: Yale University Press.

Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher, 21*(4), 12-21.

Oakes, J., et al. (1993). Creating middle schools: Technical, normative, and political considerations. *Elementary School Journal, 93*(5), 461-480.

Oakes, J., Gamoran, A., & Page, R. (1992). Curriculum differentiation, opportunities, outcomes, and meanings. In P. Jackson (Ed.), *Handbook of research on curriculum* (pp. 570-608). New York: McMillan.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal, 32*, 3-33.

Oakes, J., Ormseth, t., Robert, B., & Camp, P. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: Rand.

Oakes, J., & Wells, A. S. (1996). *Beyond the technicalities of school reform: Lessons from detracking schools*. Los Angeles: Center X, Graduate School of Education and Information Studies, UCLA.

Oakes, J., & Wells, S. (1998). Detracking for high student achievement. *Educational Leadership, 55*(6), 38-41.

Ogbu, J. U. (1974). *Social structure and cognitive behavior: A critique of the heredity-environment hypothesis and an alternative interpretation of black-white differences in IQ*: ED 097 401.

Onosko, J. J. (1990). Comparing teachers' instruction to promote students' thinking. *Journal of Curriculum Studies, 22*, 443-461.

Otto, L. B. (1975). Extracurricular activities in the educational attainment process. *Rural Sociology, 40*, 162-176.

Otto, L. B. (1976). Extracurricular activities and aspirations in the status attainment process. *Rural Sociology, 41*, 217-233.

Otto, L. B., & Alwin, D. F. (1977). Athletics, aspirations, and attainments. *Sociology of Education, 42*, 102-113.

Oxley, D. (1994). Organizing schools into small units: Alternatives to homogeneous grouping. *Phi Delta Kappan, 75*(7), 521-526.

Page, R. N. (1991). *Lower track classrooms: A curricular and cultural perspective*. New York: Teachers College Press.

Palincsar, A. S. (1986a). *Interactive cognition to promote listening comprehension.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Palincsar, A. S. (1986b). Reciprocal teaching. In A. S. Palincsar (Ed.), *Teaching reading as thinking* . Oak Brook, IL: North Central Regional Educational Laboratory.

Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh. A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities". *Journal of Learning Disabilities. v31 n4, 31*(4), 370-373.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175.

Palincsar, A. S., & Brown, A. L. (1988). Teaching and practicing thinking skills to promote comprehension in the context of group problem solving. *Remedial & Special Education (Rase), 9*(1), 53-59.

Palincsar, A. S., Brown, A. L., & Martin, S. M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist, 22*, 231-253.

Palincsar, A. S., David, Y., Winn, J., Stevens, D., & Brown, A. L. (1990, April). *Examining the differential effects of teacher- versus student-controlled activity in comprehension instruction.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston.

Palincsar, A. S., & Klenk, L. J. (1991). Dialogues promoting reading comprehension. In B. Means, C. Chelemer, & M. S. Knapp (Eds.), *Teaching advanced skills to at-risk students: Views from research and practice* (pp. 112-140). U.S. Michigan: Jossey-Bass.

Pallas, A. M., Entwisle, D. R., Alexander, K. L., & Stluka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education, 67*(1), 27-46.

Parker, L. H. (1985). *A strategy for optimizing the success of girls in mathematics: Report of a project of national significance*. Canberra: Commonwealth Schools Commission.

Parker, L. H., & Rennie, L. (1997). Teachers' perceptions of the implementation of single-sex classes in coeducational schools. *Australian Journal of Education, 41*(2), 119-133.

Pate-Bain, H., Boyd-Zaharias, J., Cain, V. A., Word, E., & Binkley, M. E. (1997). *STAR follow-up studies, 1996-1997: The student/teacher achievement ratio (STAR) project*: ED 419 593.

Patrick, H., & Townsend, M. (1995). *The influence of perceived social competence on school beginners' emergent academic intrinsic motivation.* Paper presented at the annual conference of the American Educational Research Association, San Francisco.

Patterson, L. (1991). Socio-economic status and educational attainment: A multi-dimensional and multi-level study. *Evaluation and Research in Education, 5*, 97-121.

Pena, R. A. (1997). Cultural differences and the construction of meaning: Implications for the leadership and organizational context of schools. *Education Policy Analysis Archives, 5*(10), http://epaa.asu.edu/epaa/v5n10.html.

Perret-Clermont, A.-N., Perret, J.-F., & Bell, N. (1991). The social construction of meaning and cognitive activity in elementary school children. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 41-62). Washington: American Psychological Association.

Persell, C. (1977). *Education and inequality: The roots and results of stratification in America's schools*. New York: Free Press.

Peterson, P. L. (1981). *Student diversity and student achievement in large-group and small-group approaches.* Paper presented at the meetings of the American Educational Research Association, Los Angeles.

Phelan, P., Davidson, A., & Cao, H. (1991). Students' multiple worlds: Negotiating the boundaries of family, peer, and school cultures. *Anthropology and Education Quarterly, 22*, 224-250.

Pintrich, P. R., & De Groot, E. (1990). *Quantitative and qualitative perspectives on student motivational beliefs and self-regulated learning*. Paper presented at the American Educational Research Association Conference, Boston.

Pintrich, P. R., & Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In M. L. Maeher. & P. R. Pintrich (Eds.), *Advances in motivation and achievement: Goals and self-regulatory processes* (Vol. 7, pp. 371-402). Greenwich, CT: JAI Press.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom tasks. In D. Schunk & J. Meece (Eds.), *Student perceptions in the classroom: Causes and consequences* (pp. 149-183). Hillsdale, NJ: Erlbaum.

Plank, S., et al. (1993). Effects of choice in education. In E. Rasell & R. Rothstein (Eds.), *School choice: Examining the evidence* (pp. 115-118).

Podmore, V. N. (1998). *Evaluating early childhood programmes using the strands and goals of Te Whariki, the National Early Childhood Curriculum: Final report on phases one and two to the Ministry of Education*. Wellington, NZ: New Zealand Council for Educational Research.

Podmore, V. N., & Craig, B. H. (1989). *Evaluation of exploratory studies in educational computing. Study 9, the impact of microcomputers on teachers and on the observable behaviors of children aged four to seven years: Final report*. Wellington, NZ: New Zealand Council for Educational Research.

Pong, S. (1998). The school compositional effect of single parenthood on 10th grade achievement. *Sociology of Education, 71*, 24-43.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction, 11*, 365-395.

Posner, J. K., & Vandell, D. L. (1999). After-school activities and the development of low-income urban children: A longitudinal study. *Developmental Psychology, 35*(3), 868-879.

Postlethwaite, T. N., & Ross, K. N. (1992). *Effective schools in reading: Implications for educational planners. An exploratory study.*: ED 360614.

Procidano, M. E. (1992). The nature of perceived social support: Findings of meta-analytic studies. In C. Spielberger & et al. (Eds.), *Advances in personality assessment* . Hillsdale, NJ: Lawrence Erlbaum Associates.

Putnam, R. t., Lampert, M., & Peterson, P. L. (1990). Alternative perspectives on knowing mathematics in elementary schools. *Review of Research in Education, 16*, 57-150.

Qin, Z. N., Johnson, D. W., & Johnson, R. T. (1995). Cooperative versus competitive efforts and problem solving. *Review of Educational Research, 65*(2), 129-143.

Radebaugh, B. F. (1993, October). *Democratizing educational research or why Is our nation still at risk after ten years of educational reform?*: ED 366 104.

Raphael, T. E., & McMahon, S. I. (1994). Book club: An alternative framework for reading instruction. *Reading Teacher, 48*, 102-116.

Ray, B. (1997). *Home education across the United States*. Purcellville, VA: Home School Legal Defence Association. (Available on-line: www.hslda.org/media/statsandreports/ray1997/index.stm).

Ray, B. D. (1994). *A Nationwide study of home education in Canada: Family characteristics, student achievement, and other topics*. Lethbridge: Home School Legal Defense Association of Canada.

Rehberg, R. A., & Schafer, W. E. (1968). Participation in interscholastic athletics and college expectations. *American Journal of Sociology, 73*, 732-740.

Reid, N. A., & Elley, W. (1991). *Progressive achievement tests of reading-teacher's manual*. Wellington, NZ: NZCER.

Reynolds, D. (1992). School effectiveness and school improvement: An updated review of the British literature. In D. Reynolds & P. Cuttance (Eds.), *School effectiveness: Research, policy and practice* . London: Cassell.

Reynolds, D., Sammons, S., Stoll, L., Barber, M., & Hillman, J. (1996). School effectiveness and school improvement in the United Kingdom. *School Effectiveness and School Improvement, 7*(2), 133-158.

Rice, J. K. (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational Evaluation & Policy Analysis, 21*(2), 215-229.

Rice, J. M. (1902). Education research: A test in arithmetic. *The Forum, 34*, 281-297.

Richardson, P., & Schwartz, I. (1998). *Making friends in preschool: Friendship patterns of young children with disabilities*. Ohio: Paul H. Brooks.

Riordan, C. (1985). Public and Catholic schooling: The effects of gender context policy. *American Journal of Education, 5*, 518-540.

Rist, R. (1970). Student social class and teacher expectations. *Harvard Educational Review, 40*, 411-451.

Ritter, G. W., & Boruch, R. F. (1999). The political and institutional origins of a randomized controlled trial on elementary school class size: Tennessee's project STAR. *Educational Evaluation & Policy Analysis, 21*(2), 111-123.

Robertson, D., & Symons, J. (1996). *Do peer groups matter? Peer group versus schooling effects on academic attainment*: London School of Economics, Center for Economic Performance.

Rogers, K. B. (1991). *The relationship of grouping practices to the education of the gifted and talented learner: executive summary* (Research/Technical Report 143). Washington, D.C.: National Research Center on the Gifted and Talented, Storrs, CT.

Rogoff, B. (1993). Commentary. *Human Development, 36*(1), 24-26.

Roscigno, V. (1999). The Black-White achievement gap, family-school links, and the importance of place. *Sociological Inquiry, 69*(2), 159-186.

Roscigno, V. J., & Ainsworth-Darnell, J. W. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. *Sociology of Education, In press*.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York: Wiley.

Rosenbaum, J. E. (1980). Track misperceptions and frustrated college plans: An analysis of the effects of tracks and track perceptions in the national longitudinal survey. *Sociology of Education, 53*(2), 74-88.

Rosenholtz, S. J., & Cohen, E. G. (1984). Status in the eye of the beholder. In J. Berger & J. M. Zelditch (Eds.), *Studies in expectation states theory: Pure and applied* . San Francisco: Jossey Bass.

Rosenholtz, S. J., & Simpson, C. (1984). The formation of ability conceptions: Developmental trend or social construction? *Review of Educational Research, 54*(1), 31-63.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479-530.

Rosentahl, R., & Rubin, D. B. (1978). Interperonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences, 3*, 377-415.

Rosenthal, R., & Rubin, D. B. (1982). Further meta-analytic procedures for asssessing cognitive gender differences. *Journal of Educational Psychology, 74*(5), 708-712.

Ross, J. A., & Cousins, J. B. (1995). The impact of explanation seeking on student achievement and attitudes. *Journal of Educational Research, 89*(2), 109-117.

Roswal, G. M., Mims, A. A., Evans, M. D., & Smith, B., et al. (1995). Effects of collaborative peer tutoring on urban seventh graders. *Journal of Educational Research, 88*(5), 275-279.

Rowe, K., J. (1988). Single-sex and mixed-sex classes: The effects of class type on student achievement, confidence and participation in mathematics. *Australian Journal of Education, 32*, 180-202.

Rowe, K. J., Nix, P. J., & Tepper, G. (1986). *Single-sex and mixed-sex classes: The joint effects of gender and class type on student performance in the attitudes towards mathematics.* Paper presented at the annual conference of the Australian Association for Research in Education, Melbourne.

Ruble, D. N., & Frey, K. (1991). Changing patterns of comparative behavior as skills are acquired: A functional model of self-evaluation. In S. Suls & T. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 79-113). Hillsdale, NJ: Erlbaum.

Rudner, L. M. (1998). Scholastic achievement and demographic characteristics of home school students in 1998. *Education Policy Analysis Archives, 7*(8), [Available on the World Wide Web: http://olam.ed.asu.edu/epaa/v7n8/).

Rumberger, R. W., & Willms, J. D. (1992). The impact of racial and ethnic segregation on the achievement gap in California high schools. *Educational Evaluation and Policy Analysis, 14*(4), 377-96.

Russo, R. P. (1988). *The National Education Longitudinal Study of 1988: Teacher Survey.*: ED 293 893.

Ryan, A. M., Gheen, M. H., & Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social-emotional role, and the classroom goal structure. *Journal of Educational Psychology, 90*(3), 528-535.

Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology, 89*(2), 329-341.

Rzoska, K. M., & Ward, C. (1991). The effects of cooperative and competitive learning methods on the mathematics achievement, attitudes toward school, self-concepts and friendship choices of Maori, Pakeha and Samoan children. *New Zealand Journal of Psychology, 20*, 17-24.

Sage, N. A., & Kindermann, T. A. (1999). Peer networks, behavior contingencies, and children's engagement in the classroom. *Merrill-Palmer Quarterly, 45*(1), 143-171.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective schools and effecfive departments*. London: Paul Chapman.

Sanna, L. J. (1992). Self-efficacy theory: Implications for social facilitation and social loafing. *Journal of Personality & Social Psychology, 62*(5), 774-786.

Saunders, W. M. (1989). Collaborative writing tasks and peer interaction. *International Journal of Education Research, 13*(1), 101-112.

Scardamalia, M., & Bereiter, C. (1991). Higher levels of agency for children in knowledge building: A challenge for the design of new knowledge media. *Journal of the Learning Sciences, 1*(1), 37-68.

Scardamalia, M., & Bereiter, C. (1993-1994). Computer support for knowledge-building communities. *Journal of the Learning Science, 3*(3), 265-283.

Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom into World 3. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 201-228). Cambridge, MA: MIT Press.

Schafer, W. E., & Armer, M. (1968). Athletes are not inferior students. *Trans-action, 6*, 61-62.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research, 57*(2), 149-174.

Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*, 54-61.

Schwalbe, M., & Staples, C. (1991). Gender differences in sources of self-esteem. *Social Psychology Quarterly, 54*, 158-168.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*, 321-354.

Schwartz, F. (1981). Supporting or subverting learning: peer group patterns in four tracked schools. *Anthropology and Education Quarterly, 12*, 99-121.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Seiffge-Krenke, I. (1993). Close friendship and imaginary companions in adolescence. In e. a. B. Laursen (Ed.), *Close friendships in adolescence. New directions for child development* (Vol. 60, pp. 73-87). San Francisco, CA: Jossey-Bass.

Seltzer, D. A. (1976). A descriptive study of third grade reading groups. *Dissertation Abstracts International, 36*, 09A-5811 (University Microfilms No. 76-6345).

Shanahan, T. (1999). On the effectiveness and limitations of tutoring in reading. *Review of Research in Education, 23*, 217-234.

Shanahan, T., & Barr, R. (1995). Reading recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly, 30*, 958-997.

Shanker, A. (1993). Public vs. private schools. National Forum. *Phi Kappa Phi Journal, 73*(4), 14-17.

Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of effects of class size. *American Educational Research Journal, 17*, 141-152.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. *Journal of Experimental Education, 61*(4), 293-316.

Shields, J. M. (1927). Teaching reading through ability-grouping. *Journal of Educational Methods, 7*, 7-9.

Signorella, M. L., Frieze, I. H., & Hershey, S. W. (1996). Single-sex versus mixed-sex classes and gender schemata in children and adolescents. *Psychology of Women Quarterly, 20*(4), 599-607.

Simmons, D. C., Fuchs, L. S., Fuchs, P., Mathes, P., & Hodge, J. P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning-disabled and low-performing students in regular classrooms. The Elementary School Journal, 95, 387-408. *The Elementary School Journal, 95*, 387-408.

Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton Century Crofts.

Skinner, B. G. (1953). *Science and human behavior*. New York: Free Press.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571-581.

Slavin, R. E. (1983). *Cooperative learning*. New York: Longman.

Slavin, R. E. (1984a). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(8), 6-15, 24-27.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57*, 293-336.

Slavin, R. E. (1989a). *Grouping for instruction in the elementary school*. Hillsdale, NJ: Erlbaum.

Slavin, R. E. (1989b). *School and classroom organization*. Hillsdale, NJ: Erlbaum.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*(3), 417-499.

Sletta, O., Valas, H., Skaalvik, E., & Sobstad, F. (1996). Peer relations, loneliness, and self-perceptions in school-aged children. *British Journal of Educational Psychology, 66*, 431-445.

Smith, A. B., Inder, P. M., & Ratcliff, B. (1995). The nature and context of children's conflicts. *New Zealand Journal of Educational Studies, 30*(2), 103-117.

Smith, D., & Tomlinson, S. (1989). *The school effect*: Policy Studies Institute, University of Lancaster.

Smith, K. B., & Meier, K., J. (1995). *The case against school choice*. Armonk, New York: M. E. Sharpe.

Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal, 17*, 419-433.

Sorenson, A. B., & Hallinan, M. T. (1986). Effects of ability grouping on growth in academic achievement. *American Educational Research Journal, 23*, 519-542.

Spady, W. G. (1970). Lament for the letterman: Effects of peer status and extracurricular activities on goals and achievement. *American Journal of Sociology, 75*, 680-702.

Spady, W. G. (1971). Status, achievement, and motivation in the American high school. *School Review, 79*, 379-4-3.

Spear, R. C. (1994). Teacher perceptions of ability grouping practices in middle level schools. *Research in Middle Level Education, 18*(1), 117-130.

Spreitzer, E., & Pugh, M. (1973). Interscholastic athletics and educational expectations. *Sociology of Education, 46*, 171-182.

Steedman, J. (1983). *Examination results in mixed and single sex schools: Findings from the national child development study: Report of the equal opportunities commission of the UK*. Manchester: Equal Opportunities Commission.

Steedman, J. (1984). Examination results in mixed-sex and single-sex schools. In D. Reynolds (Ed.), *Studying school effectiveness* . London: Falmer Press.

Steele, C. M. (1992, April). Race and the schooling of black Americans. *The Atlantic Monthly, 68*, 78.

Steinberg, L., Brown, B. B., Cider, M., Kaczmarek, N., & Lazzaro, C. (1988). *Noninstructional influences on high school student achievement: The contributions of parents, peers, extracurricular activities, and part time work*. Madison, WI: Clearinghouse, National Center on Effective Secondary Schools, University of Wisconsin.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Stern, P. R., & Shavelson, R. J. (1981, April). *The relationship between teachers' grouping decisions and instructional behaviors: An ethnographic study of reading instruction*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Stevens, R. J., & Slavin, R. E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal, 32*, 321-351.

Stockard, J., & Mayberry, M. (1992). *Effective educational environments*. Newbury Park, CA: Corwin Press.

Strand, S. (1998). A "value-added" analysis of the 1996 primary school performance tables. *Educational Research, 40*(2), 123-137.

Summers, A. A., & Wolfe, B. L. (1977). Do schools make a difference? *American Economic Review, 67*, 639-652.

Sweet, D. A. (1986). *Extracurricular activity participants outperform other students*. Washington, DC: United States Department of Education, OERI.

Teddlie, C., & Stringfield, S. (1993). *Schools make a difference: Lessons learned from a ten year study of school effects*. New York: Teachers College Press.

Teddlie, C., Stringfield, S., Wimpleberg, R., & Kirby, P. (1989). Contextual differences in models for effective schooling in the USA. In B. Creemer, T. Peters, & D. Reynolds (Eds.), *School effectiveness and school improvement* . Amsterdam: Swets & Zeittlinger.

The Homeschooling Federation of New Zealand. (1996). *Homeschooling in New Zealand: Some public policy issues*. Auckland, NZ: The Homeschooling Federation of New Zealand.

Thomas, A. (1998). *Educating children at home*. London: Cassell.

Thomas, S., & Mortimore, P. (1994). *Report on value-added analysis of 1993 GCSE Examination results in Lancashire*. London: Curriculum Studies Department, Institute of Education, University of London.

Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary school effectiveness. *Research Papers in Education: Policy & Practice, 11*(1), 5-33.

Thrupp, M. (1995). The school mix effect: An enduring problem in educational research, policy and practice. *British Journal of Sociology of Education, 16*(2), 183-203.

Thrupp, M. (1997a). How school mix shapes school processes: A comparative study of New Zealand schools. *New Zealand Journal of Educational Studies, 32*(1), 53-82.

Thrupp, M. (1999a). The caring role of schools: Teachers' and children's perspectives. *Children Issues, 3*(1), 15-22.

Thrupp, M. (1999b). *Schools making a difference: Let's be realistic! School mix, school effectiveness and the social limits of reform*. Buckingham: Open University Press.

Timperley, H., Robinson, V., & Bullard, T. (1999). *Strengthening Education in Mangare and Otara. Evaluation Report to the Ministry of Education, Research Division, May 1999*. Wellington, NZ: Ministry of Education.

Toma, E. F. (1996). Public funding and private schooling across countries. *Journal of Law and Economics, 39*, 121-148.

Townsend, M. A. R., & Hansen, J. B. (1986). The nature of friendships, leadership and social support among adolescent girls in social welfare home. *New Zealand Journal of Educational Studies, 21*, 158-167.

Townsend, M. A. R., & Hicks, L. (1997). Classroom goal structures, social satisfaction and the perceived value of academic tasks. *British Journal of Educational Psychology, 67*, 1-12.

Townsend, M. A. R., Manley, M., & Tuck, B. F. (1991). Academic help-seeking in intermediate school classrooms: Effects of achievement, ethnic group, sex and classroom organisation. *New Zealand Journal of Educational Studies, 25*, 35-47.

Townsend, M. A. R., McCracken, H. E., & Wilton, K., M. (1988). Popularity and intimacy as determinants of psychological well-being in adolescent friendships. *Journal of Early Adolescence, 8*, 421-436.

Trimble, K. D., & Sinclair, R. L. (1987). On the wrong track: Ability grouping and the threat to equity. *Equity and Excellence, 23*. 15-21.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *American Journal of Psychology, 9*, 507-533.

Trowbridge, D., & Durnin, R. (1984a). *A study of student-computer interactivity.* Unpublished paper, Irvine: University of California.

Trowbridge, D., & Durnin, R. (1984b). *Research from an investigation of groups working together at the computer.* Unpublished paper, Irvine: University of California.

Trussell-Cullen, A. (1994). *Whatever happened to times tables? Every parent's guide to New Zealand education.* Auckland, NZ: Reed Books.

Tuckman, B. W., & Bierman, M. (1971). *Beyond pygmalion: Galatea in the schools.* Paper presented at the annual meeting of American Educational Research Association, New York.

Tudge, J., & Rogoff, B. (1989). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In M. H. Bornstein & J. S. Brunner (Eds.), *Interaction in human development* (pp. 17-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Urberg, K. A. (1999). Introduction: Some thoughts about studying the influence of peers on children and adolescents. *Merrill-Palmer Quarterly, 45*(1), 1-12.

Urberg, K. A., Degirmencioglu, S. M., & Pilgrim, C. (1997). Close friend and group influence on adolescent cigarette smoking and alcohol use. *Developmental Psychology, 33*, 834-844.

Urdan, T., Midgley, C., & Wood, S. (1995). Special issues in reforming middle level schools. *Journal of Adolescence, 15*(1), 9-37.

Useem, E. L. (1991). Student selection into course sequences in mathematics: the impact of parental involvement and school policies. *Journal of Research in Adolescence, 1*(3), 231-250.

Useem, E. L. (1992). Middle schools and math groups: Parents' involvement in children's placement. *Sociology of Education, 65*(4), 263-279.

Valentine, J., Clark, D. D., Irvin, J. L., Keefe, J. W., & Melton, G. (1993). *Leadership in middle level education: A national survey of middle level leaders and schools.* (Vol. I). Reston, VA: National Association of Secondary School Principals. ED 356 535.

Van Fossen, B. E., Jones, J. D., & Spade, J. D. (1987). Curriculum tracking and status maintenance. *Sociology of Education, 60*, 104-122.

Vaughn, V. L. (1990). *Meta-analysis of pull-out programs in gifted education.* Paper presented at the annual convention of the National Association for Gifted Children, Little Rock, AR.

Vaughn, V. L., Feldhusen, J., & Asher, J. W. (1991). Meta-analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly, 35*(2), 92-98.

Veeman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of educational Research, 65*(4), 319-381.

Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research, 66*(3), 323-340.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge: Harvard University Press.

Wagemaker, H. (Ed.). (1993). *Achievement in reading literacy: New Zealand's performance in a national and international context.* Wellington, NZ: Research Section, Ministry of Education.

Waggoner, M., Chinn, C., Yi, H., & Anderson, R. (1995). Collaborative reasoning about stories. *Language Arts, 72*, 582-589.

Wallace, T. A. (1989). *The effects of enrichment on gifted students: A quantitative synthesis.* Unpublished doctoral dissertation, University of Illinois, Chicago.

Webb, J. (1989a). The outcomes of home-based education: Employment and other issues. *Educational Review, 41*(2), 121-133.

Webb, N. M. (1980b). Group process and learning in an interacting group. *The Quarterly Newsletter of the Laboratory of Comparative Science: Issues in Aggregation, 2*, 10-15.

Webb, N. M. (1980c). An analysis of group interaction and mathematical errors in heterogeneous ability groups. *British Journal of Educational Psychology, 50*, 266-276.

Webb, N. M. (1982a). Peer interaction and learning in cooperative small groups. *Journal of Educational Psychology, 74*, 642-655.

Webb, N. M. (1982b). Group composition, group interaction, and achievement in cooperative small groups. *Journal of Educational Psychology, 74*, 475-484.

Webb, N. M. (1984b). Sex differences in interaction and achievement in cooperative small groups. *Journal of Educational Psychology, 76*, 33-34.

Webb, N. M. (1984c). Stability of small group interaction and achievement over time. *Journal of Educational Psychology, 76*, 211-224.

Webb, N. M. (1989b). Peer interaction and learning in small groups. *International Journal of Education Research, 13*, 21-39.

Webb, N. M. (1991). Task related verbal interaction and Mathematics learning in small groups. *Journal for Research in Mathematics Education, 22*, 366-389.

Webb, N. M. (1992). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz-Lazarowitz (Ed.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 102-119). New York: Cambridge University Press.

Webb, N. M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment, 1*, 131-152.

Webb, N. M. (1995b). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 102-119). New York, NY: Cambridge University Press.

Webb, N. M. (1995c). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*(2), 239-261.

Webb, N. M., & Cullian, L. K. (1983b). Group interaction and achievement in small groups: Stability over time. *American Educational Research Journal, 20*, 411-424.

Webb, N. M., Ender, P., & Lewis, S. (1986). Problem solving strategies and group processes in small groups learning computer programming. *American Educational Research Journal, 23*, 243-261.

Webb, N. M., & Farivar, S. (1994). Promoting helping behavior in cooperative small groups in middle school mathematics. *American Educational Research Journal, 31*(2), 369-395.

Webb, N. M., & Kenderski, C. M. (1984). Student interaction and learning in small group and whole class settings. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction: Group organization and group processes* (pp. 153-170). New York: Academic Press.

Webb, N. M., & Kenderski, C. M. (1985). Gender differences in small-group interaction and achievement in high- and low-achieving classes. In L. C. Wilkinson & C. B. Marrett (Eds.), *Gender differences in classroom interaction* (pp. 209-236). New York: Academic Press.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841-873). New York, NY: Prentice Hall.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*, 406-423.

Wehlage, G. G., Rutter, R. A., Smith, G. A., Lesko, N., & Fernandez, R. R. (1989). *Reducing the risk: Schools as communities of support*. New York: Springer-Verlag.

Weinstein, R. S. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. *Journal of Educational Psychology, 68*, 103-116.

Welner, K. G., & Oakes, J. (1996). (Li)Ability grouping: the new susceptibility of school tracking systems to legal challenges. *Harvard Educational Review, 66*(3), 451-470.

Wenglinsky, H. (1997). *When money matters: How educational expenditures improve student performance and how they don't*. Princeton, NJ: The Educational Testing Service, Policy Information Center.

Wentzel, K. R. (1989). Adolescent classroom goals, standard for performance, and academic achievement: An interactionist perspective. *Journal of Educational Psychology, 81*(2), 131-142.

Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding motivation at school. *Journal of Educational Psychology, 91*, 76-97.

Wentzel, K. R., & Asher, S. R. (1995). Academic lives of neglected, rejected, popular, and controversial children. *Child Development, 66*, 754-763.

Wentzel, K. R., & Caldwell, K. (1997). Friendships, peer acceptance, and group membership: Relations to academic achievement in middle school. *Child Development, 68*, 1198-1209.

Wentzel, K. R., & Caldwell, K. (1998). Friendships, peer acceptance, and group membership: Relations to academic achievement in middle school. *Child Development, 68*, 1198-1209.

Werry, B. W. (1987). Profile of New Zealand secondary schools. In Department of Education (Ed.), *Mathematics achievement in New Zealand secondary schools: A report on the conduct in New Zealand of the second international mathematics study within the International Association for the Evaluation of Educational Achievement* (pp. 66-67). Wellington, NZ: Department of Education.

Westera, J., & Moore, D. W. (1995). Reciprocal Teaching of Reading Comprehension in a New Zealand High School. *Psychology in the Schools, 32*(3), 225-232.

Wheelock, A. (1992). *Cross the track: How "untracking" can save America's schools*. New York: New Press. ED 353 349.

Wheelock, A. (1994). *Alternatives to tracking and ability grouping*. Arlington, VA: American Association of School Administrators.

Wiatrowski, M. D., Hansell, S., Massey, C. R., & Wilson, D. L. (1982). Curriculum tracking and delinquency. *American Sociological Review, 47*, 151-160.

Wicker, A. (1968). Undermanning, performances, and students' subjective experiences in behavior settings of large and small high schools. *Journal of Personality and Social Psychology, 10*, 255-261.

Wigfield, A. (1994a). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49-78.

Wigfield, A. (1994b). The role of children's achievement values in the self-regulation of their learning outcomes. In D. Schunk & B. Zimmerman (Eds.), *Self-regulation of learning and performance* (pp. 101-124). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*, 265-310.

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem: Change across elementary and middle school. *Journal of Early Adolescence, 14*(2), 107-138.

Wigfield, A., & et al. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*(4), 552-565.

Wilce, H. (1984a). Still puzzling over the sexual equation. *Times Educational Supplement*(No. 3528, 17 February), 10.

Wilce, H. (1984b). Teachers influenced by pupils' sex and colour. *Times Educational Supplement*(No. 3528, 17 February), 10.

Wilkinson, I. A. G. (1998). Dealing with diversity: Achievement gaps in reading literacy among New Zealand students. *Reading Research Quarterly, 33*, 144-167.

Wilkinson, I. A. G., & Anderson, R. C. (1995). Sociocognitive processes in guided silent reading: A microanalysis of small-group lessons. *Reading Research Quarterly, 30*, 710-740.

Wilkinson, I. A. G., & Hamilton, R. J. (1998, December). *Teaching and learning in composite classes. Paper presented at annual meeting of the*

New Zealand Association for Research in Education. Paper presented at the annual meeting of the New Zealand Association for Research in Education, Dunedin, New Zealand.

Wilkinson, I. A. G., & Townsend, M. A. R. (in press). From Rata to Rimu: Grouping for instruction in 'best practice' New Zealand classrooms. *The Reading Teacher.*

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

Willis, S., & Kenway, J. (1986). On overcoming sexism in schooling: To marginalize or mainstream. *Australian Journal of Education, 30*, 132-149.

Willms, J. D. (1983). Do private schools produce higher levels of academic achievement? New evidence for the tuition tax credit debate. In T. James & H. M. Levin (Eds.), *Public dollars for private schools: The case of tuition tax credits* (pp. 223-231). Philadelphia: Temple University Press.

Willms, J. D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. *American Sociological Review, 51*, 221-241.

Willms, J. D. (1992). *Monitoring school performance.* London: Falmer Press.

Willms, J. D., & Raudenbusch, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects adn their stability. *Journal of Educational Measurement, 26*(3), 209-232.

Witte, J. (1996). School choice and student performance. In H. F. Ladd (Ed.), *Holding school accountable* (pp. 149-176). Washington: The Brookings Institution.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89-100.

Word, E., Johnston, J., Bain, H., Fulton, D. B., Boyd-Zaharias, J., Lintz, M. N., Achilles, C. M., Folger, J., & Breda, C. (1990). *Student/teacher achievement ratio (STAR): Tennessee's K3 class-size study.* Nashville, TN: Tennessee State Department of Education.

Wylie, C. (1996). *Five Years Old & Competent: A Summary of the Main Findings of the First Stage of the Competent Children Project for Parents and Early Childhood Education Teachers.* Wellington, NZ: New Zealand Council for Educational Research, Wellington. (ED 402 103).

Wylie, C. (1997). *Self-managing school seven years on -- What have we learnt?* Wellington, NZ: Council for Educational Research.

Wylie, C. (1998). *Can vouchers deliver better education?* Wellington, NZ: New Zealand Council for Educational Research.

Wylie, C., & Smith, L. (1995). *Learning to learn: Children's progress through the first 3 years of school.* Wellington, NZ: New Zealand Council for Educational Research.

Wylie, C., Thompson, J., & Lythe, C. (1999). *Competent children at 8 families, early education, and schools.* Wellington, NZ: New Zealand Council for Educational Research.

Youniss, J., & Smollar, J. (1989). Adolescents' interpersonal relationships in social context. In T. J. Berndt & G. Ladd (Eds.), *Peer relationships in child development* (pp. 300-316). New York: Wiley.

Zimmer, R. W., & Toma, E. F. (in press). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Influence of Peer Effects on Learning Outcomes: A Review of the Literature

Author(s): Ian A.G. Wilkinson, John A. Hattie, Judy M. Parr, Michael A. R. Townsend, Martin Thrupp, Hugh Lauder, & Tony Robinson

Corporate Source: Auckland Uniservices Limited

Publication Date: June, 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ? ☑	Level 2A ? ☐	Level 2B ? ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Ian A.G. Wilkinson
Printed Name/Position/Title: Ian A.G. Wilkinson / Dr.
Organization/Address: School of Teaching & Learning Ohio State University
Telephone: 614-292-8716
FAX: 614-292-4260
E-Mail Address: wilkinson.70@osu.edu
Date: 8-22-03

200 Ramseyer Hall
29 W. Woodruff Ave.
Columbus. OH 43210

PS 031428

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

Karen E. Smith, Acquisitions
ERIC/EECE, University of Illinois
Children's Research Center
51 Gerty Drive
Champaign, IL 61820

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to: